



STIC Search Report

Biotech-Chem Library

STIC Database Tracking Number: 179610

TO: Rosanne Kosson
Location: rem/3B84/3C70
Art Unit: 1653
Friday, February 17, 2006
Case Serial Number: 10/754115

From: Kristine Hensle
Location: Biotech-Chem Library
REM-1B69
Phone: (571) 272-4161

Kristine. Hensle@uspto.gov

Search Notes

Examiner Kosson,

See attached results.

If you have any questions about this search feel free to contact me at any time.

Thank you for using STIC search services!

Kristine Hensle
Librarian
STIC Biotech/Chem Library
(571)272-4161

179610
mej

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:45:19 ; Search time 87.0796 Seconds
(without alignments)
4606.314 Million cell updates/sec

Title: US-10-754-115-47
Perfect score: 5005
Sequence: 1 MKNIDPKLYQKTPVSVYDN.....DAEISFLTITPLKNVQPKHR 960

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5005	100.0	960	US-10-706-424-12	Sequence 12, Appl
2	5005	100.0	960	US-10-754-115-47	Sequence 47, Appl
3	2879	57.5	938	US-10-706-424-16	Sequence 16, Appl
4	2850.5	57.0	949	US-10-706-424-14	Sequence 14, Appl
5	2850.5	57.0	949	US-10-754-115-64	Sequence 64, Appl
6	2740	54.7	1043	US-10-262-794A-61	Sequence 61, Appl
7	2537.5	50.7	915	US-09-817-514A-6	Sequence 6, Appl
8	2535.5	50.7	915	US-10-754-115-58	Sequence 58, Appl
9	2304.5	46.0	962	US-10-754-115-51	Sequence 51, Appl
10	2304.5	46.0	962	US-11-020-848-4	Sequence 4, Appl
11	2226	44.5	1016	US-10-753-901-16	Sequence 16, Appl
12	2226	44.5	1016	US-10-754-115-16	Sequence 16, Appl
13	1581.5	31.6	973	US-10-754-115-61	Sequence 61, Appl
14	1581.5	31.6	930	US-10-609-113-13	Sequence 13, Appl
15	1581.5	31.6	930	US-10-754-115-42	Sequence 42, Appl
16	1581.5	31.6	953	US-10-609-113-19	Sequence 19, Appl
17	1581.5	31.6	953	US-10-754-115-43	Sequence 43, Appl
18	1268.5	25.3	940	US-10-365-742-64	Sequence 64, Appl
19	651	13.0	286	US-10-609-113-41	Sequence 41, Appl
20	339.5	6.8	2364	US-10-156-761-7834	Sequence 7834, Ap
21	332.5	6.6	932	US-10-282-122A-44617	Sequence 44617, A
22	329.5	6.6	1572	US-10-282-122A-69415	Sequence 69415, A
23	321	6.4	1530	US-10-282-122A-68175	Sequence 68175, A
24	318	6.4	1385	US-10-282-122A-68242	Sequence 68242, A
25	317	6.3	1565	US-10-282-122A-44558	Sequence 44558, A
26	315	6.3	1317	US-10-369-493-11243	Sequence 11243, A
27	307.5	6.1	843	US-10-282-122A-76214	Sequence 76214, A

ALIGNMENTS

RESULT 1

US-10-706-424-12
; Sequence 12, Application US/10706424
; Publication No. US20040103455A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens
; FILE REFERENCE: 62878
; CURRENT APPLICATION NUMBER: US/10/706,424
; CURRENT FILING DATE: 2003-11-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 12
; LENGTH: 960
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-706-424-12

Query Match 100.0%; Score 5005; DB 4; Length 960;
Best Local Similarity 100.0%; Pred. No. 1.6e-315;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MKNIDPKLYQKTPVSVYDN	GRGLIIRNIDFHR	TANGDPDTRITR	HQVDIHGLNQSIDP 60
DB	1	MKNIDPKLYQKTPVSVYDN	GRGLIIRNIDFHR	TANGDPDTRITR	HQVDIHGLNQSIDP 60
QY	61	RLYEAKQTNNTIKPNFLWQYDLTG	NPLCTESIDAGRTVTLNDIEGR	PLLTVTATGVIQTR 120	
DB	61	RLYEAKQTNNTIKPNFLWQYDLTG	NPLCTESIDAGRTVTLNDIEGR	PLLTVTATGVIQTR 120	
QY	121	QYETSSLPGRLLSVAEQTPPEEKTSR	ITERLIWAGNTEAKDHNLAGQCVR	HYDVTAGVTRL 180	
DB	121	QYETSSLPGRLLSVAEQTPPEEKTSR	ITERLIWAGNTEAKDHNLAGQCVR	HYDVTAGVTRL 180	
QY	181	ESLSLTGTVLSSQLLIDTQENW	TGNETWQNNLADDIYTTLLST	FDATGALLTQTD 240	
DB	181	ESLSLTGTVLSSQLLIDTQENW	TGNETWQNNLADDIYTTLLST	FDATGALLTQTD 240	
QY	241	KGNIQRLAYDVAGQLNGSWLTKG	QTEQVIKSLTYSAAQKLR	EEHGNVDVTEYSYEPE 300	
DB	241	KGNIQRLAYDVAGQLNGSWLTKG	QTEQVIKSLTYSAAQKLR	EEHGNVDVTEYSYEPE 300	
QY	301	TORLIGIKTRPSDKTKVLODL	RYEYDYPGVNVISIRNDAEAT	EFWNQKMPENTYD 360	
DB	301	TORLIGIKTRPSDKTKVLODL	RYEYDYPGVNVISIRNDAEAT	EFWNQKMPENTYD 360	
QY	361	YOLISATGREMANIQSQSHOR	PPSPALPSDNNNTYNTYTRY	TYDTRGNLTQKHSSPATON 420	
DB	361	YOLISATGREMANIQSQSHOR	PPSPALPSDNNNTYNTYTRY	TYDTRGNLTQKHSSPATON 420	

QY 421 NYTNTIVTSNRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480
 Db 421 NYTNTIVTSNRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480
 QY 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTORVTVLPLNLELRLTQNSTATTEDL 540
 Db 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTORVTVLPLNLELRLTQNSTATTEDL 540
 QY 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600
 Db 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600
 QY 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660
 Db 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660
 QY 661 LYRMVRNNPVTLDDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720
 Db 661 LYRMVRNNPVTLDDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720
 QY 721 LPKASTSSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSTTST 780
 Db 721 LPKASTSSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSTTST 780
 QY 781 NLOKKSFTLYRADNRSFEEMQSKPEPKANTPLDTKWAROPASIFIGOKDTSNLPKETV 840
 Db 781 NLOKKSFTLYRADNRSFEEMQSKPEPKANTPLDTKWAROPASIFIGOKDTSNLPKETV 840
 QY 841 KNIISTWGAAPKLKOLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
 Db 841 KNIISTWGAAPKLKOLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
 QY 901 GOKNPLPEGRTKNMVPSLLDTPQIETSSIIALNHGPNVDABISFLLTPIPLKNVPHKR 960
 Db 901 GOKNPLPEGRTKNMVPSLLDTPQIETSSIIALNHGPNVDABISFLLTPIPLKNVPHKR 960

RESULT 2
 US-10-754-115-47
 ; Sequence 47, Application US/10754115
 ; Publication No. US20040208907A1

GENERAL INFORMATION:

APPLICANT: Hey, Timothy
 APPLICANT: Schleper, Amanda
 APPLICANT: Bevan, Scott
 APPLICANT: Bintrim, Scott
 APPLICANT: Mitchell, Jon
 APPLICANT: Li, Ze Sheng
 APPLICANT: Ni, Weiting
 APPLICANT: Zhu, Baolong
 APPLICANT: Merlo, Don
 APPLICANT: Apel-Birkhold, Patricia
 APPLICANT: Meade, Thomas
 TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
 FILE REFERENCE: DAS-104XC1
 CURRENT APPLICATION NUMBER: US/10/754,115
 CURRENT FILING DATE: 2004-01-07
 PRIOR APPLICATION NUMBER: US 60/441,723
 PRIOR FILING DATE: 2003-01-21
 NUMBER OF SEQ ID NOS: 64
 SOFTWARE: PatentIn version 3.2
 SEQ ID NO 47
 LENGTH: 960
 TYPE: PRT
 ORGANISM: Photorhabdus luminescens
 US-10-754-115-47

Query Match 100.0%; Score 5005; DB 4; Length 960;
 Best Local Similarity 100.0%; Pred. No. 1.6e-315;
 Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHRITANGDPDTRITRHOYDIHGHLNQSIDP 60
 QY 61 RLYEAKQTNNTIKNFNLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATATGVIQTR 120
 Db 61 RLYEAKQTNNTIKNFNLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATATGVIQTR 120
 QY 121 QYETSSLPGRLLSVAEOTPEEKTSTRITELRLWAGNTEAEKDHNLAGQCVRHVDYTAGVTRL 180
 Db 121 QYETSSLPGRLLSVAEOTPEEKTSTRITELRLWAGNTEAEKDHNLAGQCVRHVDYTAGVTRL 180
 QY 181 ELSLGTGTVLSQSSQLLIDTQEANWTGDNETVWQNLADDIYTTLTSTFDATGALLTQTD 240
 Db 181 ELSLGTGTVLSQSSQLLIDTQEANWTGDNETVWQNLADDIYTTLTSTFDATGALLTQTD 240
 QY 241 KGNITQRLAYDVAGOLNGSWLTKGQTEQVIIKSLTYSAGQKLREHNDVITEYSYEP 300
 Db 241 KGNITQRLAYDVAGOLNGSWLTKGQTEQVIIKSLTYSAGQKLREHNDVITEYSYEP 300
 QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWHNQKVPENTTYD 360
 Db 301 TORLIGIKTRRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWHNQKVPENTTYD 360
 QY 361 YQLISATGREMANIGQOSHQPSPALPSDNNNTYNTYTRTYTYDRGNLTKIQHSSPATON 420
 Db 361 YQLISATGREMANIGQOSHQPSPALPSDNNNTYNTYTRTYTYDRGNLTKIQHSSPATON 420
 QY 421 NYTNTIVTSNRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480
 Db 421 NYTNTIVTSNRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOQVTLV 480
 QY 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTORVTVLPLNLELRLTQNSTATTEDL 540
 Db 481 KRDKGANDREWYRSGDGRMLKINEQQAASNAQTORVTVLPLNLELRLTQNSTATTEDL 540
 QY 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600
 Db 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEYYYP 600
 QY 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660
 Db 601 YGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYYPQWIGRWLSSDPAGTIDGLN 660
 QY 661 LYRMVRNNPVTLDDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720
 Db 661 LYRMVRNNPVTLDDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAINRPPVAPKPS 720
 QY 721 LPKASTSSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSTTST 780
 Db 721 LPKASTSSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSTTST 780
 QY 781 NLOKKSFTLYRADNRSFEEMQSKPEPKANTPLDTKWAROPASIFIGOKDTSNLPKETV 840
 Db 781 NLOKKSFTLYRADNRSFEEMQSKPEPKANTPLDTKWAROPASIFIGOKDTSNLPKETV 840
 QY 841 KNIISTWGAAPKLKOLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
 Db 841 KNIISTWGAAPKLKOLSNVIKTKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
 QY 901 GOKNPLPEGRTKNMVPSLLDTPQIETSSIIALNHGPNVDABISFLLTPIPLKNVPHKR 960
 Db 901 GOKNPLPEGRTKNMVPSLLDTPQIETSSIIALNHGPNVDABISFLLTPIPLKNVPHKR 960

RESULT 3

US-10-706-424-16
 ; Sequence 16, Application US/10706424
 ; Publication No. US20040103455A1
 GENERAL INFORMATION:
 APPLICANT: ffrench-Constant, Richard
 APPLICANT: Waterfield, Nicholas
 TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens
 FILE REFERENCE: 62878

Query Match	57.0%;	Score 2850.5;	DB 4;	Length 949;
Best Local Similarity	60.9%;	Pred. No. 8.1e-176;		
Matches 583;	Conservative 100;	Mismatches 207;	Indels 67;	Gaps 17
Qy	1	MKNIDPKLYOKTPTVSYVDNRGLIIRNIDFHRITANGDPDTRIIRHQYDIHGLHMQSIDP	60	
Db	1	MKNIDPKLYOHTPTVNVYDNRGLTIRNIDFHRDVAGGDTDFIRHQYDTRGHLSQSIDP	60	
Qy	61	RLYEAKQTNNTI1KPNFLWQYDLTGNPLCTESIDAGRVTYTLNDIEGRPLTTLVTATGVQTR	120	
Db	61	RLYDAKQTNNSNPFLWQYINLTGDTLRTESVDAGRVALNDIEGRQVLIVTATGAIQTR	120	
Qy	121	OYETSSIPUGRLLSVAEQTPPEKTSRITERLLWAGNTEAEKHNLAGQCVRRHYDVTAGVTRL	180	
Db	121	OYBANTUPUGRLLSVSEQAPGEQTPRVTEHFWAGNTOAEKHNLAGQVRRHYDVTAGVTQL	180	
Qy	181	ESLSLTGTVLSQSQSLI1DQEAENWTGNETVWQNMLADDIYTTLTSTFDATGALLTQDDA	240	
Db	181	ESLSLTENILSQSQALLADQEAEDMTGNDETLMTQKLNSEYTTQGSTFDATGALLTQDDA	240	
Qy	241	KGM1ORLAYDVAGOLGNSWLT1LKQOTQVQI1KSLTYSAAGQKLREHGNVDVITEYSYEP	300	
Db	241	KGM1ORLAYNVAGOLGNSWLT1LKQSEQVI1KSLTYSAAGQKLREHGNVITEYSYEP	300	
Qy	301	TQRLIGIKTRRPSDKVLQDLRYEYDVPGVNVISIRNDAEATFRFHNQKVPENTVITYDSL	360	
Db	301	TLRLIGITTRQSDKVLQDLRYEHDVPVGNIIISVRNDAEATFRFHNQKIPENTVITYDSL	360	
Qy	361	YQLISATGREMANIGQOSHOPSP--ALPDSNNYTYNTRYTYTYDRGGNLTAKI0HSSPAT	418	
Db	361	YQLISATGREMANIGQSQNLQPSPI1PLPTDENSYNTYTRSYNDRGGLVQIRHSSPAA	420	
Qy	419	QNNYTNNTIVSNRSNRNAVLSLT1LTDPAQVDALFDAGGHQNTLI1SQNLNWNTRGELQV	478	
Db	421	QNNYTTDITVSNRSNRNAVLSLTSDPTQVEALFDAGGHQTKLLPGQELSWNTRGELQV	480	
Qy	479	LVKDKGANDDREWYRYSGBGRRLMKNEQQASNNAAQTRVYTL1PNLELR1LTONSTATTE	538	
Db	481	PVSRESAS--DREWYR1GNDGWRRLKVSSEQTGSTQOORVYTL1PDLRL1TQNGTTTSE	538	
Qy	539	DLQVITVGEAGRAQVRVLHWESGKPEID1NNQLRYSYDNLI1GSSOLELDSQEQI1SEREY	598	
Db	539	DLHAI1TVGAAGHAQVRVLHWETT1PPAG1NNQLRYSYDNLI1GSSOLELDSQEQI1SOBEY	598	

NUMBER OF SEQUENCES: 88
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dow AgroSciences Patent Department
 STREET: 9330 Zionsville Road
 CITY: Indianapolis
 STATE: IN
 COUNTRY: US
 ZIP: 46268
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/262,794A
 FILING DATE: 02-OCT-2002
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/951,567
 FILING DATE: 05-MAY-1997
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/063,615
 FILING DATE: 18-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/395,497
 FILING DATE: 28-FEB-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/007,255
 FILING DATE: 06-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/608,423
 FILING DATE: 28-FEB-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/705,484
 FILING DATE: 28-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J
 REGISTRATION NUMBER: 27386
 REFERENCE/DOCKET NUMBER: 960296.93804
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9166
 INFORMATION FOR SEQ ID NO: 61:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1043 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-10-262-794A-61

Query Match 54.7%; Score 2740; DB 4; Length 1043;
 Best Local Similarity 74.7%; Pred. No. 1.4e-168;
 Matches 513; Conservative 78; Mismatches 88; Indels 8; Gaps 3;
 1 MKNIDPKLYOKTPTVSVYVNRGLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSIDP 60
 1 MSPSETLLYTPTVSVLDNRGLSIRDIGFHRVIGGDTKTRVTRHOYDARGHLNYSIDP 60
 61 RLYEAKQTNNTIKNFVLDNRGLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSIDP 120
 61 RLYDAKQADNSKFNFWQHLAHLALTESVDAGRTVALNDIEGRSVMTWNTATGVRQTR 120
 121 QYETSSLPGRLLSVAEQTPPEKTSRI TERLIWAGNTEAEKHNLACQCVHYDTAGVTRL 180
 121 RYEGNTLPGRLLSVAEQTPPEKTSRI TERLIWAGNTEAEKHNLACQCVHYDTAGVTRL 180
 181 ELSLITGVLSQSSQLLIDTQEANWTDGNETVWQNMALADDIYTLTSTPDATGALLTQTD 240
 181 MSQSLAGAMLSSQHLAEGQEAQNSGDDTETVWQMLASEVYTTQSTTNAIGALLTQTD 240
 241 KGNITQRLAYDVAGQINGSLWLTKGTEQVILKSLTYSAGOKLREHNDVITEYSYPE 300

Db 241 KGNITQRLAYDVAGQINGSLWLTKGTEQVILKSLTYSAGOKLREHNDVITEYSYPE 300
 QY 301 TORLIGIKTRR-PSDTKVLQDLRYEYDPVGNVISIRNDAAEATRFWHNQKVPENTYTT 356
 Db 301 TORLIGITTRAEQSQSGARVLQDLRYKYDPVGNVISIHNDAAEATRFWRNKVPEENRYV 360
 QY 357 YDSLYQLISATGREMANIGQOSHOPSPAL--PSDNNTYTYTTRYTYTDRGGLTKIHS 414
 Db 361 YDSLYQLMSATGREMANIGQSNQLPSPVIEPVTDDSTYTYLRYTYTYTDRGGLVQIRHS 420
 QY 415 SPATONNYTNNITVSNRRAVLSTLTEDPAQVDALPDAGGHONTLISQNLNWNTRGEL 474
 Db 421 SPATQNSYTTDITVSSRNRRAVLSTLTEDPAQVDALPDAGGHONTLISQNLNWNTRGEL 480
 QY 475 QQVTLVKRDKGANDREWYRYSGDGRRLKINEQQAASNAQTRVYLPNLELRLTQNST 534
 Db 481 QRVTVPVSRN--SSDSEWYRYSSDGMRLKVKSEQQTGNSQTQVRVYLPGLRLTTGVAD 538
 QY 535 ATTEDLQVITVGEAGRAQVRVLHWESKGPEDIDNNQLRYSYDNLIGSQLELDSGQIIS 594
 Db 539 KTTEDLQVITVGEAGRAQVRVLHWESKGPEDIDNNQVRYSYDNLIGSQLELDSGQIIS 598
 QY 595 EEEYYPYGGTALWAARNQTRVYSGKGERDATGLYYGYRYVYOPWIGRWLSSDDPAG 654
 Db 599 QEEYYPYGGTALWAARNQTRVYSGKGERDATGLYYGYRYVYOPWIGRWLSSDDPAG 658
 QY 655 TIDGLNLYRMVRNPNVTLDPGLMPT 681
 Db 659 TVDGLNLYRMVRNPNVTLDPGLMPT 685
 RESULT 7
 US-09-817-514A-6
 ; Sequence 6, Application US/09817514A
 ; Patent No. US20020078478A1
 ; GENERAL INFORMATION:
 ; APPLICANT: French-Constant, Richard
 ; APPLICANT: Bowen, David
 ; APPLICANT: Rocheleau, Thomas
 ; APPLICANT: Waterfield, Nicholas
 ; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
 ; FILE REFERENCE: 61645
 ; CURRENT APPLICATION NUMBER: US/09/817,514A
 ; CURRENT FILING DATE: 2000-03-26
 ; PRIOR APPLICATION NUMBER: US 60/191806
 ; PRIOR FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 915
 ; TYPE: PRT
 ; ORGANISM: Photorhabdus luminescens
 US-09-817-514A-6

Query Match 50.7%; Score 2537.5; DB 3; Length 915;
 Best Local Similarity 70.1%; Pred. No. 1.5e-155;
 Matches 488; Conservative 75; Mismatches 108; Indels 25; Gaps 6;
 1 MKNIDPKLYOKTPTVSVYVNRGLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSIDP 60
 1 MSSYNSAIDQKTPSIKVLNDRKLVRLTYLEURTQADNSDELITFYBFIPIFGVKSDP 60
 61 RLYEAKQTNNTIKNFVLDNRGLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSIDP 120
 61 R-----KNKQSGNFIRFVNLACQVLRSESDAGRTITLNDIESRPVLIINATGVRQNH 115
 121 QYETSSLPGRLLSVAEQTPPEKTSRI TERLIWAGNTEAEKHNLACQCVHYDTAGVTRL 179
 116 RYEDNTLPGRLLAITEQVAGEKT---TERLIWAGNTPQEKDYNLACQCVHYDTAGLTQ 172
 180 LESLSLITGVLSQSSQLLIDTQEANWTDGNETVWQNMALADDIYTLTSTPDATGALLTQTD 239
 173 LNSLSAGVLSQSSQLLIDTQADWTDGNETVWQNMALADDIYTLTSTPDATGALLTQTD 232

QY	240	AKGNIQRLAYDVAGQLNGSWLTLXGQTEQVVKSLTYSAAQOKLREEHGNDVITEYSEP	299
DB	233	AKGNIQRLAYDVAGQLNGSWLTLXGAQEQVVKSLTYSAAQOKLREEHGNGIVITEYSEP	292
QY	300	ETORLIGIKTRRPSDTKVLQDLRYEYDVGWVISIRNDAEATRFWNGQKVPENTYTVDS	359
DB	293	ETORLIGIKTRRPSDAKVLQDLRYQYDVGWVINIRNDAEATRFWRNGKVPENSYTVDS	352
QY	360	LYQLISATGREMANIGQOSHQFPSPALPDSNNYTNVTRTYTYDRGGLTKIOHSSPATQ	419
DB	353	LYQLISATGREMANIGQONNQLPSPALPSDNNYTNVTRSYSDHSGNLTKIRHSSPATQ	412
QY	420	NNYTYTNVTSNRRAVLSTLTPDPAQVDALFDAGCHQNTLISGQNLNWNTRGELQQVTL	479
DB	413	NNYTYTNVTSNRNRGVLSLTLTPDNOVDTLFDAGGHQTSLLPGQTLTWTPRGELKQV--	470
QY	480	VKRDKGANDREWTRYSGDGRMLUKINFGQASNNAAQTORVYVLPNLEURLTQNSTATTED	539
DB	471	---NNGPGN--EWYRYSNNGMRQLKVSEQPTNTTQQQRVYVLPGLERLTTQSNATTTT	525
QY	540	LQVITVGEAGAOVRVLHWESGKPEDIDNNOLRYSYDNLIGSSQLELDSGQIISEEYV	599
DB	526	LHVITLGEAGAOVRVLHWESGKPEDVNNQLRYSYDNLIGSSQLELONQGQIISEEYV	585
QY	600	PYGGTALWAARNQTEASYKTYRYSKGERDATGLYYGYRYTQPWIGRWLSSDPACTIDGL	659
DB	586	PFGGTALWAANSQTEASYKTYRYSKGERDATGLYYGYRYTQPWAGRWLSDPACTIDGL	645
QY	660	NLYRMVRNPNFVTLDPDGLMPTIAERIAALKKNKVTT	695
DB	646	NLYRMVRNPNVSLQDENG LAP-----BKGYT	672

RESULTS

```

US-10-754-115-58
; Sequence 58, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 58
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-754-115-58

```

Db	61	R-----KXKQSGNFIRVFNLAGOVREESVDAGRITLNDIESRPPVLIINATGVRQNH	115
Qy	121	QYETSPLPGRLLSVAEQTPPE--EKTISRITERLIWAGNTEAEKXHDNLNLAGOCVRHYDTAGVTR	179
Db	116	RYEDNTLPGRLLAITEQVQAGEKT--TERLIWAGNTPQEKDYNLAGOCVRHYDTAGLTQ	172
Qy	180	LESLSLTGTVLSQSQQLLIDTOEANWTDGNETVWQNMLADDIYYTLSTFDATGALLTQTD	239
Db	173	LNSLSLAGVLSQSQQLLVDVDKNADTGEDQSLWQKQLSSDVYTTQNKADATGALLTQTD	232
Qy	240	AKGNIQRLAYDVAGQLNGSWLTLKGQTEQVTKSLTYSAAQOKLREEHGNDVITEYSYEP	299
Db	233	AKGNIQRLAYDVAGQLKGCWLT.LKGAEQVTKSLTYSAAQOKLREEHGNGVITEYSYEP	292
Qy	300	ETORLIGIKTRRPSDTKVQLDLRYEYDPVGNVISIRNDAEATRFWNNOKVPENYTYTDS	359
Db	293	ETORLIGIATRPSDAKVQLDLRYQYDPVGNVINIRNDAEATRFWRNOKVPENSYTYDS	352
Qy	360	LYQLISATGREMANIGQSHQFPSPALPSSDNNYTYNTRYTYTYDRGMLTKIQHSSPATQ	419
Db	353	LYQLISATGREMANIGQNNQLPSPALPSSDNNYTYNTRYSYDHSGLNLTQIRHSSPATQ	412
Qy	420	NNYTTNITVNSRNRVLSTLTEDPAQVDALFDAGGHQNTLISGQNLNWRTEGELQQVTL	479
Db	413	NNYTTVAITLSNRNRGLSTLTITDPNQVDTLFDAGGHQTSLLPGQTLTWTPEGLKQV--	470
Qy	480	VKRDKGANDDREWRYSGDGRMLKINBQQAASNNAAQORVTYLPNLRLTLNSTATTED	539
Db	471	---NNGPGN--EWTRYDSNGNRQLKVSQPTQNTTQQORVLYLPGLELRTQSNATTTEE	525
Qy	540	LQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELSEGOIISSEEEYY	599
Db	526	LHVITLGEAGRAQVRVLHWESGKPEDVNNQLRYSYDNLIGSSQLELNDQGOIISSEEEYY	585
Qy	600	PYGCTALWAARNQTEASVKTIRYSGKERDATGLYYGYRYQPTWGRWLSLSPAGTIDGL	659
Db	586	PFGGTALWAANSQTEASVKTIRYSGKERDATGLYYGYRYQPTWAGRWLSADPAGTIDGL	645
Qy	660	NLYRMVRNNPVTLTDPDGLMPTIAERIAALKKQKVT	695
Db	646	NLYRMVRNNPVSLQDENG LAP-----EKGKVT	672

RESIT.T 9

```

US-10-754-115-51
; Sequence 51, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 51
; LENGTH: 962
; TYPE: PRT
; ORGANISM: Xenorhabdus bovienii
US-10-754-115-51

```


Query Match 46.0%; Score 2304.5; DB 4; Length 962;
 Best Local Similarity 64.6%; Pred. No. 2.1e-140;
 Matches 445; Conservative 94; Mismatches 145; Indels 5; Gaps 4;

QY 1 MKNIDPKLYOKTPTVVVDNRGLIIRNIDFHRITTAGNDPDRITRPHQYDIHGLNQSDP 60
 DB 1 MNVFNPTLYAGTPTVTVMNDRGLSVRDIAHYHRTTAGEQADTRITRHOQYSPHNFLESIDP 60

QY 61 RLVEAKQTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIOTR 120
 DB 61 RLFDL-QSOSTIKPNFTYCPALKGDLVLTESVDAGQTVILSDIEGRPLLNISAMGVVGH 119

QY 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGAGQCVRRHYDTAGVTRL 180
 DB 120 QYESTLPGRLLAVERKNEASTPQIIERFIWSGNSPSEKDHNLGAGLYRHYDTAGLNOL 179

QY 181 ELSLGTGTVLSQSSQLLIDTQENWNTGNETVQNMALADDIYTLSTFDATGALLTQDA 240
 DB 180 NAVSLTSDLSQSRQLLQDDVTADWSGDSQWKTRLSNDIFTTBITADAVGNFLTQDA 239

QY 241 KGNTRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHNDVITEYSYEPE 300
 DB 240 KSNQRLSYDVAGQLKASWLTIKGNEQVIVNSLTYSAGOKLREBQGVVTEYSYEAQ 299

QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDYPVGNVISIRNDAEATRFWMNQKMPENTTYDLSL 360
 DB 300 TWRLIGVTAYRQSKKRLQDLVYNDPVGNLIRNNAEATRFWRNQIVEPENHYAYDSL 359

QY 361 YOLISATGREMANTQOQSHQPPSP--ALPSDNNTVNTYTRTYHYDRGNLTKIQHSSPAT 418
 DB 360 YOLISASGREIASIQOQSRLPVPPIPLPANDDVVTRYTRYHYDRGNLTCQIRHCAPAT 419

QY 419 QNNYTNITVNSNRNAVLTITLTPAQVDALFDAGGHQNTLISGNNMNTNRGELQOVT 478
 DB 420 DNKYTKITVNSNRNAVWLTITLTPAKVDTLFDHGHQLQLOSGQTLQWYRGELQOIT 479

QY 479 LVKRDKGANDREWRYSGDGRMLKINEQASNNAAQTVRYTLPNLELRTQNSTATTE 538
 DB 480 KIQRDE-KPADKERYRYGVGAARVVKISTQOAGGSSHVQVRVYLPGLRLTQHDATLIE 538

QY 539 DLQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLGSSQLELDSGQIIISEREY 598
 DB 539 DLQVIIMGEAGRAQVRVLHWEIPPPDNLNDSRLSYSDLSMGSSQLELDGAGQIITQBEY 598

QY 599 YPYGCTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYPWIGRWLSSDPAGTIDG 658
 DB 659 LNLRYMVRNPNVTLDPDGLMPTIAERIA 687

QY 659 LNLRYMVRNPNVTLDPDGLMPTIAERIA 687
 DB 659 LNLRYMVRNPNVTLDPDGLMPTIAERIA 686

RESULT 10
 US-11-020-848-4
 ; Sequence 4, Application US/11020848
 ; Publication No. US20050155104A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Hey, Timothy
 ; APPLICANT: Hey, Timothy
 ; APPLICANT: Sheets, Joel
 ; APPLICANT: Meade, Tom
 ; APPLICANT: Li, Ze Sheng
 ; APPLICANT: Lira, Justin
 ; APPLICANT: Russell, Sean
 ; APPLICANT: Thompson, Robin
 ; APPLICANT: Mitchell, Jon
 ; APPLICANT: Fencil, Kristin
 ; TITLE OF INVENTION: Toxin Complex Proteins and Genes from Xenorhabdus bovienii
 ; FILE REFERENCE: DAS-114XCI
 ; CURRENT APPLICATION NUMBER: US/11/020,848
 ; CURRENT FILING DATE: 2004-12-23

; PRIOR APPLICATION NUMBER: US 60/534,893
 ; PRIOR FILING DATE: 2004-01-07
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 4
 ; LENGTH: 962
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus bovienii
 US-11-020-848-4

Query Match 46.0%; Score 2304.5; DB 6; Length 962;
 Best Local Similarity 64.6%; Pred. No. 2.1e-140;
 Matches 445; Conservative 94; Mismatches 145; Indels 5; Gaps 4;

QY 1 MKNIDPKLYOKTPTVVVDNRGLIIRNIDFHRITTAGNDPDRITRPHQYDIHGLNQSDP 60
 DB 1 MNVFNPTLYAGTPTVTVMNDRGLSVRDIAHYHRTTAGEQADTRITRHOQYSPHNFLESIDP 60

QY 61 RLVEAKQTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIOTR 120
 DB 61 RLFDL-QSOSTIKPNFTYCPALKGDLVLTESVDAGQTVILSDIEGRPLLNISAMGVVGH 119

QY 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGAGQCVRRHYDTAGVTRL 180
 DB 120 QYESTLPGRLLAVERKNEASTPQIIERFIWSGNSPSEKDHNLGAGLYRHYDTAGLNOL 179

QY 181 ELSLGTGTVLSQSSQLLIDTQENWNTGNETVQNMALADDIYTLSTFDATGALLTQDA 240
 DB 180 NAVSLTSDLSQSRQLLQDDVTADWSGDSQWKTRLSNDIFTTBITADAVGNFLTQDA 239

QY 241 KGNTRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHNDVITEYSYEPE 300
 DB 240 KSNQRLSYDVAGQLKASWLTIKGNEQVIVNSLTYSAGOKLREBQGVVTEYSYEAQ 299

QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDYPVGNVISIRNDAEATRFWMNQKMPENTTYDLSL 360
 DB 300 TWRLIGVTAYRQSKKRLQDLVYNDPVGNLIRNNAEATRFWRNQIVEPENHYAYDSL 359

QY 361 YOLISATGREMANTQOQSHQPPSP--ALPSDNNTVNTYTRTYHYDRGNLTKIQHSSPAT 418
 DB 360 YOLISASGREIASIQOQSRLPVPPIPLPANDDVVTRYTRYHYDRGNLTCQIRHCAPAT 419

QY 419 QNNYTNITVNSNRNAVLTITLTPAQVDALFDAGGHQNTLISGNNMNTNRGELQOVT 478
 DB 420 DNKYTKITVNSNRNAVWLTITLTPAKVDTLFDHGHQLQLOSGQTLQWYRGELQOIT 479

QY 479 LVKRDKGANDREWRYSGDGRMLKINEQASNNAAQTVRYTLPNLELRTQNSTATTE 538
 DB 480 KIQRDE-KPADKERYRYGVGAARVVKISTQOAGGSSHVQVRVYLPGLRLTQHDATLIE 538

QY 539 DLQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLGSSQLELDSGQIIISEREY 598
 DB 539 DLQVIIMGEAGRAQVRVLHWEIPPPDNLNDSRLSYSDLSMGSSQLELDGAGQIITQBEY 598

QY 599 YPYGCTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYPWIGRWLSSDPAGTIDG 658
 DB 599 YPYGCTALWAARNQTEANYKTIIRYSGKERDATGLYYGHRYYPWIGRWLSSADPAGTVDG 658

QY 659 LNLRYMVRNPNVTLDPDGLMPTIAERIA 687
 DB 659 LNLRYMVRNPNVTLDPDGLMPTIAERIA 686

RESULT 11
 US-10-753-901-16
 ; Sequence 16, Application US/10753901
 ; Publication No. US2004019416A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bintrim, Scott
 ; APPLICANT: Mitchell, Jon
 ; APPLICANT: Larrinua, Ignacio
 ; APPLICANT: Apel-Birkhold, Patricia
 ; APPLICANT: Schafer, Barry

Db 590 SDGQIIISQEEYYPYGGTAVMAARSQSSEADYKTVRYSGKERDATGLYYYGYRYYSQSWTGRW 649
Qy 648 LSSDPAGTIDGLNLYRMVRNPNVTLDDPDGLMP 680
Db 650 LSVDPAGEVDGLNLFMRCPNPIVPSDSGGRFP 682

RESULT 13
US-10-754-115-61
; Sequence 61, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 61
; LENGTH: 973
; TYPE: PRT
; ORGANISM: Serratia entomophila
US-10-754-115-61

Query Match 43.6%; Score 2184; DB 4; Length 973;
Best Local Similarity 60.5%; Pred. No. 1.4e-132;
Matches 435; Conservative 87; Mismatches 149; Indels 48; Gaps 9;

Qy 8 LYQKTPVSVVDNRGLIIRNIDFHRITANGDPDT-----RITHQYDIHGLNQSIDPR 61
Db 5 LFSSTPVAVLDNRGLLRELYR-----HPDTPETDERITCQHDGSLGSLQSADPR 59
Qy 62 LYEAQTNNTIKPNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATG---VI 117
Db 60 LHAAGLT-----NFTYLSNLTGTVLQSVSDAGTSLSDAAGRAFLAVTGAGTEDAVT 113
Qy 118 QTRQYETSSLRGLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLAGQCVRHYDTAGV 177
Db 114 RTWQYEDDTLPGRLPSITEQVTGE-AAQITERFVYAGNTDAEKILNLAGQCVSHYDTAGL 172
Qy 178 TRLESLSITGTVLSSOSSQLLDTQEAQNTGNETWQNMADDIYTLTSTFDATGALLTQ 237
Db 173 VQTDLSIALSGVPLAVTRQLLPAAGANWNGDASAWNDLLDGETFFQTQHADATGAVLSI 232
Qy 238 TDAGNIORLAYDVAGQLNGSWLTKGQTEOVIIKSLTYSAGQKLRREHGNVDVITYSY 297
Db 233 TDAGNLRQVAYDVAGLLSGSWLTKGQTEQIVIVASLYSAGKLRREHGNVGVTSIY 292
Qy 298 EPETQRLIGIKTRRPS-----DTKVLQDLRIYEDVPVGNVISIRNDAEATRFWNNQKMPEN 353
Db 293 EPETQRLTGIKTERPSGHVAGAKVLQDLRIYEDVPVGNVLSVNNDAEATRFWRNQNKPEN 352
Qy 354 TYTDSLSQLISATGREMANTGQSHQPPSPA--LPSDNNTYNTYTRYTYDRGNLTKI 411
Db 353 TYIYDLSLQVLSATGREMANAGQGNLPSATAPLPTDSSATYNTYTRYTYDRGNLTQM 412
Qy 412 QHSSPATONNTYNTYTRYTYDRGNLTSTLTEDPAQVDALPDAGGHONTLISGONLNWNT 471
Db 413 RHPAPATNNNTYTDITVSDRSNRNRAVLSTLAEVPSVDMLFSGAGHQKHLQPGQALVWTPR 472

Qy 472 GELQOVLVKRDKGANDREWRYSGGRRMLKINEQQAASNAQTQRTVYILPNLELRITQ 531
Db 473 GELQKVTVPVRDGA--DDSESYRYDAGSQRIIKTGTQRTGNVQTRVVYILPGLERIMA 531
Qy 532 NSTATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSVDNLLGSSQLELDSGQ 591
Db 532 NGVTEKESLQVITVGEAGRAQVRVLHWEIGKPDLDDEDSVRYSDNLLVGSSQLELDREGY 591
Qy 592 IISBEYYPYGGTALWAARNQTEASYKTIRYSGKERDATGLYYYGYRYYPWIGRWLSSD 651
Db 592 LISBEFYYPYGGTAVLTARSEVADYKTIRYSGKERDATGLDYGYRYYPWAGRWLSTD 651
Qy 652 PAGTIDGLNLYRMVRNPNVTLDDPDGLMPT-----IAERIAALKK 691
Db 652 PAGTVDGLNLFMRVRNPNVTLFDSNGRISTQGEARRLVGEAFVHPLHMPVFERISVERK 710

RESULT 14

US-10-609-113-13
; Sequence 13, Application US/10609113
; Publication No. US20040110184A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Bevan, Scott
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from
; TITLE OF INVENTION: Paenibacillus Species
; FILE REFERENCE: DAS-101XC2
; CURRENT APPLICATION NUMBER: US/10/609,113
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US 60/392,633
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/441,647
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 13
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Paenibacillus strain IDAS 1529
US-10-609-113-13

Query Match 31.6%; Score 1581.5; DB 4; Length 930;
Best Local Similarity 42.2%; Pred. No. 1.5e-93;
Matches 370; Conservative 128; Mismatches 284; Indels 95; Gaps 23;

Qy 8 LYQKTPVSVVDNRGLIIRNIDFHRITANGDPDTRITEHQYDIHGLNQSIDPRLYEAKQ 67
Db 6 IYRGTPISVVVDNRNLEIRILOYNRIAAEDPADECILRNTYTPLSILGSSMDPRLFQVQ 65
Qy 68 TNNTIKPNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATGVIQTRQYETSSL 127
Db 66 DDRCPTPNIRTWASLRGELCSVDAGRAKELFDIEGRPVWLIDANGTETTLEYD---V 122
Qy 128 PGRLLSVAEQPPEKTSRITERLIWAGNTEAE-KHNLAGQCVRHYDTAGVTRLESLSLT 186
Db 123 LGRPTAVFEQGEQDTSQCRRERFIY-GEKEADAQANRLRGQLVRHYDTAGRIQTDLSILA 181
Qy 187 GTVLSSQSLLDITQE-ANWTGDMETWQNMADDIYTLTSTFDATGALLTQTDAGNIQ 245
Db 182 GLPLRQRQLLKNWDEPDGMSMBESAWASLLAAEAYDTSWRYDAQDRVLAQTDAGNIQ 241
Qy 246 RLAYDVAGQLNGSWLTKGQTEQIVIIKSLTYSAGQKLRREHGNVDVITYSEYEPETORLI 305
Db 242 QLTNDAGQPOAVSLKQCGQAEQRIWNRIEYNAAGQVDLAEGAGIVITYEYEESTORLI 301
Qy 306 GIKTRR---PSDTKVLQDLRIYEDVPVGNVISIRNDAEATRFWNNQKMPENITYYDLSYQ 362
Db 302 RKDSRGLSSGEREVLDQRYEYDVPVGNILSIYNEAEPVRYFRNQAVAPKQRAYDALYQ 361
Qy 363 LISATGREMANIGQSHQFP--SPALPSDNNTYNTYTRYTYDRGNLTQKHSSPATON 420

```
Db 362 LVSSSGRESDALRQOTSPLLTP-IPLDSDQYVYNAEKSYDQAGNLIKLSHNG---AS 417
Qy 421 NYTNTITVSNRNRNRAVLSTLTFDPAQVDALFDAGGHQNTLLSGQNLNWNTRGELQOQVTLV 480
Db 418 QYTTNVYVYDSSNRGIWRO-GEDIPDIAASFDRAGNQOALFPGRPLEWDTNRQLSRVHMV 476
Qy 481 KRDKGANDREWYRSGDGRMLKINEQOASNNQTORVTVLPNLELRLTQNSTATTEDL 540
Db 477 VRE-GGDNDWEGYLDSSGMRIVKRSTRKTQTTQTTDTTLLPGLLELRIRQTGDRVTEAL 535
Qy 541 QVITVGEAGRAQVRVLHWESG-KPEDIDNNQLRYSYDNLIGSSOLELDSGQIIESEYY 599
Db 536 QVITVDE-GAQVRVLHWEDGTEPGGIANDQYRYSLNDHLTSSLEVDGQGIISKEEY 594
Qy 600 PYGKTALWAARNQTEASVKTIRYSGKERDATGLYYGYRYYQPIWIGRWLSSDPAGTIDGL 659
Db 595 PYGKTALWTARSEVEASYKTIRYSGKERDATGLYYGHRYYMPWLGRWLNPDPAGMVDGL 654
Qy 660 NLYRMVRNPNVTLDDPGLMP-TIA-----BRIAALK---KNKVTDSAPS 700
Db 655 NLYRMVRNPNPGLMDPFGNAPINVADYSFVHGDLVYGLSKERGRYLKLFNPNFNNEKSIDS 714
Qy 701 PANA-----TNVAINI-----RPP--VAPKPSLPKASTSSQPTTHP-----I 735
Db 715 PAMVIDQYNNVALSITNQYKVEELMKFQDPQKAARKIKVPEGNLSRNNENYPLWHDYI 774
Qy 736 GAANIKPTSGSSIVAPL-----SPVGNKSTSEISLSPESAQS8SSS 776
Db 775 NIGEAKAFAKASHIFQEVKNGYKDYHKLKLLDRMIESPLMKRGSKLGLBIAA----- 828
Qy 777 TTSTNLQKK-SFTLYRADNRSEFEQSKFPGFKAWT 812
Db 829 ---TNQRTKIHFLV---DNLNIEQVTVKEGSGGQSIT 859
```

RESULT 15

```
US-10-754-115-42
; Sequence 42, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrin, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 42
; LENGTH: 930
; TYPE: PRF
; ORGANISM: Paenibacillus strain DAS1529
US-10-754-115-42
```

```
Query Match 31.6%; Score 1581.5; DB 4; Length 930;
Best Local Similarity 42.2%; Pred. No. 1.5e-93;
Matches 370; Conservative 128; Mismatches 284; Indels 95; Gaps 23;

Qy 8 LYQKTPVSVYDNRGLIIRNIDFHRHTTANGDPDTRITHQYDINGHLNQSIDPRLYEAKQ 67
Db 6 IYRGTPITSVYDNRNLEIRILQYNRAAEDPADECILENTYTPLSYLGSSMDPRLFSQYQ 65
```

Search completed: February 16, 2006, 21:54:20
Job time : 90.0796 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:43:14 ; Search time 21.6251 Seconds
(without alignments)
3670.210 Million cell updates/sec

Title: US-10-754-115-47
Perfect score: 5005
Sequence: 1 MKNIDPKLYQKTPVSYVDN.....DAEISFLTTIPLKNVPHKR 960

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA.*

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/pCTUS COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4545	90.8	954	2	US-09-251-645-12
2	2740	54.7	1043	2	US-08-851-567B-61
3	2537.5	50.7	915	2	US-09-817-514A-6
4	329	6.6	1627	2	US-09-328-352-6604
5	322.5	6.4	998	2	US-09-252-991A-28424
6	303.5	6.1	1439	2	US-09-543-681A-7560
7	300.5	6.0	1426	2	US-09-492-709A-340
8	291	5.8	1586	2	US-09-543-681A-5329
9	291	5.8	1526	2	US-09-252-991A-23805
10	288.5	5.8	1696	2	US-09-328-352-5542
11	287.5	5.7	974	2	US-09-252-991A-23640
12	277	5.5	1377	2	US-09-711-164-467
13	261.5	5.2	1665	2	US-09-543-681A-4476
14	245	4.9	1584	2	US-09-251-645-6
15	193.5	3.9	1128	2	US-09-252-991A-13032
16	186	3.7	804	2	US-09-328-352-5545
17	179	3.6	2736	2	US-09-252-991A-30227
18	178.5	3.6	10182	2	US-09-134-001C-3159
19	178	3.6	408	2	US-09-543-681A-5372
20	170	3.4	344	2	US-09-328-352-6657
21	167.5	3.3	1015	2	US-09-134-000C-6204
22	167.5	3.3	1306	2	US-09-538-092-330
23	165.5	3.3	2137	2	US-09-134-001C-4463
24	164.5	3.3	5024	2	US-09-710-279-2964
25	164	3.3	2283	2	US-10-172-502-4
26	162.5	3.2	1394	2	US-08-296-791-2
27	162.5	3.2	1394	2	US-09-839-996-2

28	162.5	3.2	1394	2	US-10-080-505-2	Sequence 2, Appli
29	162.5	3.2	1394	2	US-10-645-655-2	Sequence 2, Appli
30	162.5	3.2	1394	4	PCT-US95-10661A-2	Sequence 2, Appli
31	161	3.2	1395	2	US-10-080-505-7	Sequence 7, Appli
32	160.5	3.2	3290	2	US-09-328-352-5486	Sequence 5486, Ap
33	159	3.2	1187	2	US-09-949-016-6513	Sequence 6513, Ap
34	159	3.2	2169	2	US-09-949-016-6930	Sequence 6930, Ap
35	158	3.2	1537	1	US-08-325-267A-2	Sequence 2, Appli
36	157	3.1	1335	2	US-09-134-001C-3716	Sequence 3716, Ap
37	157	3.1	1848	2	US-08-296-791-6	Sequence 6, Appli
38	157	3.1	1848	2	US-09-839-996-6	Sequence 6, Appli
39	157	3.1	1848	2	US-10-080-505-6	Sequence 6, Appli
40	157	3.1	1848	2	US-10-645-655-6	Sequence 6, Appli
41	157	3.1	3696	4	PCT-US95-10661A-6	Sequence 6, Appli
42	157	3.1	1439	2	US-09-134-001C-5080	Sequence 5080, Ap
43	156.5	3.1	1439	2	US-09-134-000C-6133	Sequence 6133, Ap
44	156.5	3.1	1589	2	US-09-543-681A-4998	Sequence 4998, Ap
45	155	3.1	2353	2	US-09-377-155-33	Sequence 33, Appli

ALIGNMENTS

RESULT 1

US-09-251-645-12
; Sequence 12, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; TITLE OF INVENTION: AND NUCLEIC ACID SEQUENCES CODING THEREFOR
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251.645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 954
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-12

Query Match	90.8%	Score 4545;	DB 2;	Length 954;
Best Local Similarity	90.4%	Pred. NO. 0;		
Matches 868;	Conservative 40;	Mismatches 46;	Indels 6;	Gaps 2;
QY	1	MKNIDPKLYQKTPVSYVDNRLGILIRNIDFRTTANGDPDTRITRHQYDIHGLNQSIDP	60	
DB	1	MKNIDPKLYQKTPVSYVDNRLGILIRNIDFRTTANGDPDTRITRHQYDIHGLNQSIDP	60	
QY	61	RLYEAKQNTNTIKFNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR	120	
DB	61	RLYEAKQ-----KSNFLWQYDLTGPNLCTESVDAGRTVTLNDIEGRPLLTATGVIQTR	115	
QY	121	QYETSSLPGRLLSVAEQPEKTSRITRLLIWAQTEAKDHNLAGQCVRHVDYTAGVTRL	180	
DB	116	QYETSSLPGRLLSVTEQIP-EKTSRITRLLIWAQTEAKDHNLAGQCVRHVDYTAGVTRL	174	
QY	181	ESLSITGTVLSQSSQLLDTQEAANWGTGNETVWQNLADDIYTLTSTFDATGALLTQIDA	240	
DB	175	ESLSITGTVLSQSSQLLDTQEAANWGTGNETVWQNLADDIYTLTSTFDATGALLTQIDA	234	
QY	241	KGNIQRLAYDVAGQLNGSWLTKGTEQVIIKSLTYSAGQKLREHGNVDITEYSYBPE	300	
DB	235	KGNIQRLAYDVAGQLNGSWLTKGTEQVIIKSLTYSAGQKLREHGNVDITEYSYBPE	294	
QY	301	TORLIGIKTRPRSDTKVLQDLRYEYDVPVGNVISINDAEATRFVHKNQKMPENTYDLSL	360	

[illegible]

Qy	1	MGNIDPKLYQKPTTVSVVDNRGLIRINIDFRIITANGDPDTRIPEHQYVDIEGHLNQSDP	60
Db	1	MSPESTLYTQTPTVSVLDNRGLSIRIDIGFIRIVIGGDTDRVTRHQYDARGHLNYSIDP	60
Qy	61	RLYBAKOTNTNIKFNFLWQVLDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR	120
Db	61	RLYDAKQADNSVKFNFWQHDLAGHALRTESVDAGRTVALNDIEGRSVMTNATGVRQTR	120
Qy	121	QYETSSIPGRLLSVAOPEPKTSIRITERLIWAGNTEAEKDHNLACQCVRHYDRTAGVTRL	180
Db	121	RYEGNTLPGRLLSVSEQVFNQESAKATVERFIWAGNTTSEKYNLSGCLRHYDRTAGVTRL	180
Qy	181	ESLSLTGTVLSSQSLDLIDTOEANWTKGNETVWQNLADDTLTTLSTFDATGALLTQDQA	240
Db	181	MSQSLAGAMLSSQHLLLAEGQANWSGDETFWQGLLASEVYTTQSTTNALGALLTQDQA	240
Qy	241	KGNLQRLAYDVAGQLNGSWLTKGQTEQVITKSLTYSAAQOKLREEHGNVITEYSIPE	300
Db	241	KGNLQRLAYDIAGQLKGSWLTVKQSEQVIVKLSLSWSAAGHKLREEHGNVITEYSIPE	300
Qy	301	TQRLIGTKTRR - - - PSDTKVQLDLRYEYDVPGVNISIRNDAEATRFWHNKQMPENTYT	356
Db	301	TQRLIGITTRRAEGSQSGARVLQDLRYKYDVPGVNISITHNDAEATRFWRNKQVBPENRYV	360
Qy	357	YDSLXYQLISATGREWANIQQOSHQPSPAL - PSDNNNTYTNVTRYTYDRCGNLTKIOHS	414
Db	361	YDSLXYQLMSATGREWANIQQOSNQLPSPVIVPFDSDSYTYTWRLRYTYDRCGNLVOIRHS	420
Qy	415	SPATQNNYTNITVSNRSNRNVLSTLTEDPAQVDALFDAGGHQNTLISGONLNNWTRGEL	474

```

Db 421 SPATQSYTTDITVSSRSRAVLSTLTTPDTRVDALFDSGGHQKMLIPQNLDMNIRGEL 480
Qy 475 QOVLTVKRDKGANDREWYRSGDGRMLKINEQQAASNNAAQORVYVLPNLELRLTQNST 534
Db 481 QVTPVPSREN--SSSEWYRYSDDGMWLLKXVEEQQTGSTQVRYVLPGLRLTTGVAD 538
Qy 535 ATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIS 594
Db 539 KTTEDLOVITVGEAGRAQVRVLHWSGKPTDIDNNQVRYSYDNLIGSSQLELDSGQIIS 598
Qy 595 BEEYYPYGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSDDPAG 654
Db 599 QBEYYPYGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSADPAG 658
Qy 655 TIDGMLNLYMRVNRNPNVTLDDPGLMPT 681
Db 659 TVDGLNLYMRVNRNPNVTLDDHGLAPS 685

```

RESULT 3

US-09-817-514A-6
 ; Sequence 6, Application US/09817514A
 ; Patent No. 6639129
 ; GENERAL INFORMATION:
 ; APPLICANT: French-Constant, Richard
 ; APPLICANT: Bowen, David
 ; APPLICANT: Rocheleau, Thomas
 ; APPLICANT: Waterfield, Nicholas
 ; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
 ; FILE REFERENCE: 61645
 ; CURRENT APPLICATION NUMBER: US/09/817,514A
 ; PRIOR FILING DATE: 2000-03-26
 ; PRIOR APPLICATION NUMBER: US 60/191806
 ; FILING DATE: 2000-03-24
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 6
 ; LENGTH: 915
 ; TYPE: PRT
 ; ORGANISM: Photorhabdus luminescens

US-09-817-514A-6

Query Match 50.7%; Score 2537.5; DB 2; Length 915;
 Best Local Similarity 70.1%; Pred. No. 2.8e-182;
 Matches 488; Conservative 75; Mismatches 108; Indels 25; Gaps 6;

```

Qy 1 MNIDPKLYOKTPTVSVVDNRGLIIRNIDFHTTANGDPDTRITHQYDIHGLNQSIDP 60
Db 1 MSSYNSAIDOKTPTSIKVLNDRKLNVRTLEYLRQTQADENSDELITFYEFNIFQVVKSTD 60
Qy 61 RLYEAKQTNNTIKPNFLQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120
Db 61 R-----KNNQSGPNFIRVFNLAGQVLRBSVDAGRTITLNDIESRPVLIINATQVRQNH 115
Qy 121 QYETSSLPRLISVAEQTPE-EKTSRITERLIWAGNTEAEKHNLAGOCVRHYDTAGVTR 179
Db 116 RYEDWTLFGRLLAITEQVQAGEKT--TERLIWAGNTPQEKDYNLAGOCVRHYDTAGTQ 172
Qy 180 LESLSTGTVLSSQQLLIDTQEANWTGNETWQWMLADLIYTLTSTFDATGALLTQTD 239
Db 173 LNSLSLAGVLSQSQQLLTDQADWTDGDSLMQOKLSSDYITQSTNDATGALLTQTD 232
Qy 240 AKGNLQRLAYDVAGOLGNSWLTAKGTQVITKSLTYSAAQKGLREHGNVDVITEYSYEP 299
Db 233 AKGNLQRLAYDVAGOLKGSWLTAKQAQVITKSLTYSAAQKGLREHGNGLVITEYSYEP 292
Qy 300 ETQRLIGIKTRRPSDKVLQRLRYEYDPVGNVISIRNDAEATRFWNKQWMPENTYTYDS 359
Db 293 ETQRLIGIKTRRPSDAKVLQRLRYQYDPVGNVINRDAEATRFWNKQVAPENSYYTDS 352
Qy 360 LYQLISATGREMANIGQSHQPPSPALPSPDNNTYNTYTRSYSDHSGNLTQIRHSSPATQ 419

```

```

Db 353 LYQLISATGREMANIGQNNQLPSPALPSPDNNTYNTYTRSYSDHSGNLTQIRHSSPATQ 412
Qy 420 NNYTNTITVSNRRAVLSTLTEDPAQVDALFDSGGHQKMLIPQNLDMNIRGELQOVL 479
Db 413 NNYTVAITLNSNRNGVLSLTTPDQVDTLFDAGHOTSLPGQTLTWTPRGELKQV-- 470
Qy 480 VKRDKGANDREWYRSGDGRMLKINEQQAASNNAAQORVYVLPNLELRLTQNSTATTED 539
Db 471 ---NNGPGN--EWYRYSNGMRQLKVSQPTQNTTQQORVYVLPGLRLTTQSNATTTEE 525
Qy 540 LOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEY 599
Db 536 LHVITLGEAGRAQVRVLHWSGKPEDVNNQLRYSYDNLIGSSQLELDSGQIISSEY 585
Qy 600 PYGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSDDPAGTIDGL 659
Db 586 PFGGTALWAANSQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGWLSADPAGTIDGL 645
Qy 660 NLYMRVNRNPNVTLDDPGLMPTIAERIAALKKQKVT 695
Db 646 NLYMRVNRNPNVSLQDENGAP-----BKGYT 672

```

RESULT 4

US-09-328-352-6604
 ; Sequence 6604, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 6604
 ; LENGTH: 1627
 ; TYPE: PRT
 ; ORGANISM: Acinetobacter baumannii

US-09-328-352-6604

Query Match 6.6%; Score 329; DB 2; Length 1627;
 Best Local Similarity 22.2%; Pred. No. 2e-15;
 Matches 197; Conservative 105; Mismatches 293; Indels 294; Gaps 40;

```

Qy 14 TVSVVDNRGLIIRNIDFHTTANGDPDTRITH-----TTQDDGATSYFEYDEKQNLTCMVDAEQGRWPKQYDGSNL 54
Db 696 TSYTYDERGNVL-----TTQDDGATSYFEYDEKQNLTCMVDAEQGRWPKQYDGSNL 748
Qy 55 NQSIDPRLYEAKQTNNTIKPNFLQY-DLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTA 113
Db 749 IKEIDPLKHETAYVYNAM--GLVTSITDAKGSKSLKYDDQGNLISYTDGSGK-----799
Qy 114 TGVQTR-QYETSSLPRLLSV-----AEQTPPEKTSRI-----TERL 150
Db 800 ----ETKWQYDER---GRVISIENALNQKVEFYFTELLENREPIIKGLPLNAFQLEKI 852
Qy 151 IWAGNTEAEKHNLAGOCVRHYD-TAGVTRLESLSLTGTVLSSQSL-----196
Db 853 KHADGTEHFHDAEGRLLAHVDPKQNTRYE-YDEAGLILSRDALNHLKLYKWDRLGR 911
Qy 197 ---LIDTQEANWT-----GDNETVWQ-NMLADDIYTLTSTFDATGALLTQ 237
Db 912 LTRLINENGASYQFYDVASRLVKBEIDPDGKETVYHYDEKSQLATSIEVASAYQDLKD 971
Qy 238 TDA-KGNLQRLAYDVAGOLN-----GSLWTLKQTEQV-----IKSLT-----275
Db 972 RAAPKDRITQPIFDSMGRLEQRTAGYGHYGLEEEK-QTEEFAYDYMGRIIQAQNAQSNL 1030
Qy 276 ---YSAAGQKLRHGHND-----VITEYSVEPETQRLIGIKTRRPSDKVLQDLRYEYD 326
Db 1031 QWIFYAAG-NLVQHEHQDYKINKTAVWKHQYDEINDR---IKTRP-DGQVIDWLTGSG 1085

```



```
Db 1248 NPLQIDPWGLLT-----WDDARSGACTGVC----- 1275
Qy 728 SQPHTHPGAAIKPTTSGSSIVAPLSPGVGNKSTSEISLPSAQSS-----SS 775
Db 1276 -----GVLRIIGSKFSDTADALD--KETQNRSLCNDMEYSGIVCKDTNGKYPAS 1327
Qy 776 STTSTNLOKKSFTLYR 791
Db 1328 KAETDNLKRESYPLKR 1343

RESULT 8
US-09-543-681A-5329
; Sequence 5329, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 5329
; LENGTH: 1586
; TYPE: PRT
; ORGANISM: Proteus mirabilis
; NAME/KEY: UNSURE
; LOCATION: (1576)
; FEATURE:
; OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-543-681A-5329

Query Match 5.8%; Score 291; DB 2; Length 1586;
Best Local Similarity 21.5%; Pred. No. 1.4e-12;
Matches 209; Conservative 121; Mismatches 358; Indels 282; Gaps 43;
Qy 18 YDNRGLIIR-----NIDFHTTANGDPDTRITRHQVDIHGLNQSIDPRLYE 64
Db 704 YDERGNLTQWTDPEQOSTCYDWDKDFAPPTAQTLPNGAAWHWEYNEHGDIRRVDPGLHI 763
Qy 65 AKQTNNTIKPNFLQYDITGNPLCTESIDA-----GRVTVLNDIEGRP-LITV 111
Db 764 TR-----LAWD-----DOGLCLGQVDKGNETHYRNARGQLIEQDCSGYPTTLTY 810
Qy 112 TATGVIOT---RQVETS---SLPGRLLSVAEQTPPEKTSR-----ITERLIWAGNTEAEK 160
Db 811 DDWQRLSLTWAQNETTYTTFSEAGLLT--ECLPDGTENYDYDATGQLV--GITDAGE 866
Qy 161 DH-----NLACQCVR-----HYDTAGVTRLESLSLTGTVLSOSSQLLIDTOBAN 204
Db 867 RHILLRRNRGQVIARRDPAGHWLHFHYDTFG--RMQAL----- 903
Qy 205 WTGNETWQNLADDIYTTISTPDATGALLTQTDKAGNIORLAYDVAGQLNGSWLTKG 264
Db 904 ---ENEQGEQYRF-----EYDALHRLTDEHDLIGQQKHQYQVDMGNVTQIKTTGPG 951
Qy 265 QTEQVI-----IKSLTYSAAQKREEHGNDVITEYSPEPTQRLIGI-----KTR 310
Db 952 SIDTPIPLSPQVTFGDKVGRLLIFRENA-DYRTEYLYOPLSVLRRVPMVWHEAERTG 1010
Qy 311 RPSDTKVLQDLRYEYDPGVNVI-----SIRND----- 337
Db 1011 TTARVEYQDALTFYTDKVGQLVREASARGDYQHHDYVLNGLNITRITELPHORAFEYLYXGSG 1070
Qy 338 -AEATRFHWNQKWPENTYTTDSLYQLISAT-----GREWANTGOQ--SHOF 381
Db 1071 HLQQTQWRDNEQLTVLAEYQYDRDLRHRETLRTSGALDNETGYDCRQTHQVARQNASQF 1130
Qy 382 PSPALP-----SDNNTVTN---YTRTYVDRGGNLTKIQHSSPATQ 419
```

```
Db 1131 VTPVIDRRYRWDKRNQLIERSVSGTGEVFTAGHWYVHSYQYDPLQQLTAHLGVSQTEH 1190
Qy 420 NNYTNTITVNSRNRVAVLSTLTEDPAQVDALFDAGG-----HQNTLISQNLNWNTRGEL 474
Db 1191 FLYDAANALLTRPTEAPHNQVQSDKFDYRYDGFGRMVSRYERKSSSGQRYHYDSD--- 1247
Qy 475 QQVTLVKDKG-ANDREWRYSGDGRMLKINQOASNNNAQTORVTYLPNLELRLTQNS 533
Db 1248 HRIITAVDIDQPLGYQRAEYRYDILGRRIEK-RLWKASAIANT--VYVQH----- 1295
Qy 534 TATTEDLQVITVGEAGRAQV-----RVLHW---ESCKP-----EDIDN-----NOL 571
Db 1296 ---EPDEVYTFGWGMRLVSEHSAAPHHTTVYHAYNDQSVYTPLARIECTDNPLNPORAI 1351
Qy 572 RYSYDNLIGSQLELDSGQIISBEEYYPYCGTALWAARNOQTEASYKTIRVSGKERD-AT 630
Db 1352 YYTHSSLSGLPEALTNSEGEIVWGOQVSAWHLQQRTPRTSTFNREQNLRFGQYFDKET 1411
Qy 631 GLTYGYRYYOPMIGRWLSSDPACTIDGLNLYRMVRNNPVTLLDPDGLMPTIAERIAALK 690
Db 1412 GLHNTFRYAPDLGRFTQDDPIGLAGGINLIYAY-PNPLTWDPFGW--SCGSDAKILR 1468
Qy 691 KKVYTDAPSANATNVAINIRPPVAPKPSLPKASTSSQPTTHPIGAANIKPTTSGSSIV 750
Db 1469 ENMIERGIIIP-NYKNSA-----HHIIMSNAXDVRM-RWLR 1502
Qy 751 APLSPVG---NKSTSEISLPSAQSSSTSTTNLQKKSFT-LYR-----ADNRS 796
Db 1503 RKNRLGIDINDAINGVLTSSKIKIDVGTSTTAHSHKIHTALYKQNVFDRLNKIKDATE 1562
Qy 797 FEEMQSKFPE 806
Db 1563 FROELINKIAE 1572

RESULT 9
US-09-252-991A-23805
; Sequence 23805, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 23805
; LENGTH: 1626
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-23805

Query Match 5.8%; Score 291; DB 2; Length 1626;
Best Local Similarity 21.2%; Pred. No. 1.5e-12;
Matches 167; Conservative 95; Mismatches 284; Indels 242; Gaps 32;
Qy 10 QKTTVSV-----YDNRGLIIRNIDFHTTANGDPDTRITRHQVDIHGLNQSIDP-----R 61
Db 863 EQLPDGVSVRYRYDTLG-----RQITRQDESGAITRFQWDAAGRLSQITLPGGASR 913
Qy 62 LYAEKQTNNTIKPNFLWOYDLTGNPLCTESIDAGRTVTNLNDIEGRPLIT--VTATGVIQT 119
Db 914 MY-----RYNAYGK-VTSECDEQGRMTRYEYVLDLHLVSRREINPDGSQLR 957
Qy 120 RQYETSSLPGRLLSVAGQTPPEKTSRITERLIWAGNTEAEKHNLAGOCVHR-YDTAGVT 178
Db 958 YRYENARL---LLSEIENERGER-----YRLDYHGNGLISQETGDFGRRTAYRYDLKG-- 1007
```


Qy	179	RLESLSLTGTVLSSOQLLIDTQBANWTGDNETVWQNMLADDIYTTLLSTFDATCALLTOT	238
Dd	1008	-----QLEKTEYGGDGSSELRTTYQR-----DSTGRLLAKT	1038
Qy	239	DAKNIORLAYDVAGOL-----NGSW-LTLKGOTEQVIKS-----LYTSAAQKL--	283
Dd	1039	LPGDNRYDYRDTLGRLVAVDDGTWPPLAYEYDLDRILVRHQGWATLHYDAYDALGOQLHC	1098
Qy	284	-----REEHGNDIVETYSYPETORLI	305
Dd	1099	RLPCGNRVRYQTGGTLSAIDLNGQSLTRHFQSGREROQQGELLISQTHYDEOGRLLA	1158
Qy	306	GIKTRPSDTKVLODLREYDPGVNVISIRNDABEATRFWNQKVPENTTYDSLYQLIS	365
Dd	1159	HQVSQR---QRHLYQRYVDASGNLAALIEDSRKGRSFH-----YDPLDRLLG	1204
Qy	366	ATGR-----EMANTGOOSHQPSPALP-SDNTVTNTTRYTYDRGNGNLTKIQHSSP	416
Dd	1205	VRGETPESFVHPDPAIGNLACQGQDFARQMVEVRGNRLLTQCDRIFFDYDAHGNVLRRERGTC	1264
Qy	417	--ATONNYTNITVSNRSNRNAVLTSLTEPPAQVDALFDAGGHQ-NTLISQGLNWNNTRG	472
Dd	1265	QKLUTEYSYD-----CQHRLLGVSLPDGRQVVYRYDAFGRIAQXOVGRWTEFWLWG	1316
Qy	473	ELOQVTLVGRDKGANDREW--YRISGDGRMLKINEQQASNNNAQTORVTPLPNLELRLT	530
Dd	1317	E-----RLLAESGDRHYRTYLYPEDSPRPALLD-----	1345
Qy	531	QNSTATTEDLQVITVGZBAGRAQVRVLHWESCKPBEDIINNQLRYSYDNLIGSSQLELDSEG	590
Dd	1346	-----GEQ-----PEQVE--PCYQLDHL-GTPOELTRADG	1373
Qy	591	QIIISEEBYYPYGCTALWAARNQOTSEASYKTIRYSCKERDA-TGLYYGYRYYPQWIGRWLS	649
Dd	1374	RLCMSARYAYGN-----VLKLDIAEVNPNURFOQYFYDAETGLHYNRHRYNPSTGYRLT	1429
Qy	650	SDPAGTTIDGGLNLYMRVRNPNVTLLDPDGLMPTTAERIAALKKCNKVTDSAPSPANATNVAI	709
Dd	1430	PDPRLAGGLNSYRYV-PNPTGWVDVPLGLASNKTRCPREALDNN-----PRKGINSTS	1480
Qy	710	NIRPPVAP	717
Dd	1481	GSKKPAGP	1488

```

RESULT 10
US-09-328-352-5542
; Sequence 5542, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5542
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5542

Query Match      5.8%; Score 288.5; DB 2; Length 1596;
Best Local Similarity 22.2%; Pred. No. 2.2e-12;
Matches 179; Conservative 123; Mismatches 316; Indels 189; Gaps 39;

Qy      19  DNRGLIIRNIDFH-RTTANG-----DPDTRITRHOYDIHGHLNQSIDP-----60
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      637  DKKGKITRQIDFDGRETQOEYNDQDLVKIVQPNGGIIIRPAYKNQGNLVEIKDPEGSIWK 696
           ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy      61  RLY-EAKQTNNTIKP-NFLWQYDLTGNPCLTESIDAG---RTVTLLNDIEGRPLLTVTATG 115

```

Db	697	REYDENRNVKBNPLGHIYQYKNNNDQVLVDKAGGVKKIQYNEL-QOMLSYTDCSG	755
Qy	116	VICTORYETSSLPERLISVAEQPEEKTSRI-----TERLIWAGNTBAEKDHNL	164
Db	756	KSSTWEYDED--GAL--TAEQTANNKVQVFYSTKGRDKGQLQSIIYPDLKEYFEHDE	810
Qy	165	AGQCVRHYPAGVTRLESLSLTGTVLSQSOLLIDQEAN-----WTGDNETWQNMLAD	219
Db	811	EGRLKHDTDKGLVTEYKYNQV-----LAEQRIDANRHSVAYQMDKQGRIOKLINQ	862
Qy	220	DIVTTLSTFDATGALLTODTDAKNIORLAYDVAGOL-----NGSWLTLKGQTE-----QVI	270
Db	863	NOAEYLGYNPYGVLIREQAFDEEKHYSNENGRLFQIRPNILT---QFYIYADQOIA	919
Qy	271	IKSLTYSAAQKULREEHG-----NDVITYSEYEPBT---QRLIGIKTRRPSDTKVLQ----	319
Db	920	SKSFTHLTGOKOTEQPDYNLNSQLSRASNEVSQIDLYRNALGQLVREHGYKIPELKPL	979
Qy	320	--DLREYDVPGVNI--SIRNDAEATRE-----WHNOKVMPENTYTYDSLYQLI	364
Db	980	TAVLHYEYDELGNIKTIIRPDGHTLNLHYGSGHIVAIGLNNQEV---SFQRDDLHR--	1034
Qy	365	SATGREMAN-----IGQOSHOPFPSPALPFDNNNTYNTT--RTYTYDRGGLTKIQ	412
Db	1035	-ETTRLLANGLMTKQYNDVGLLSSQFNQP--EQETQDYLQYQAHRKHYDKNYLLSOVE	1091
Qy	413	HS-----SP-----ATQNNYTNITVNSRNSRAVLSTLTEDPAQVDALFDAGGHQ	458
Db	1092	DSRLGKLNQYDPIGRIILAAQSLHKTESFNFDPAGNLIDSESVLSPAQI-----KN	1142
Qy	459	TLI---SGQNLNWNTRGELQOQVTLVKRDKGANDREW-----YRYSGD	498
Db	1143	NLIKSYGKGYHDVQGVTEII-----QAGKNLKLTDWNQNRILRSNNGNLVTEYGYDVF	1198
Qy	499	GRMLKINEQOAS-----NNAQ-----TORVTVLPNLEURLTQNSTATTEDLQ	541
Db	1199	GRRLYKKTAKRELTLFGWDGLMTIWESFKAQNTVTKHYIYEPDSFVPLLQ--AGYKDFIQ	1256
Qy	542	VITVGEAGRAQVR-----VLHWESGKPEDIDNNQLRYSYDNLIGSSOLELDSSEQII	593
Db	1257	LIEPTDPQBYOTKPYSIYKDPVWNRNLGK-ERTALBQFTFCHCQCVGTPQTMNIRECV	1315
Qy	594	SEEEYYPYGCTALWAARNQTEA-SYKTI RYSGKERD-ATGLYTYGYRYQFWMIGRWLSSD	651
Db	1316	WEILOQDTWAGVSOIKALNQDNPEQNNLRFQGYDYRETELHYNRYRYEYEPHSARYVSKD	1375
Qy	652	PAGTIDGLNLYRMVRNPNVTLDPDGL	678
Db	1376	PIGLEGGNTSSYV-SDPNQWIDPKGL	1401

```

; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5542
; LENGTH: 1596
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5542

```

```

Query Match      5.8%; Score 288.5; DB 2; Length 1596;
Best Local Similarity 22.2%; Pred. No. 2.2e-12;
Matches 179; Conservative 123; Mismatches 316; Indels 189; Gaps 39
Qy 19 DNRGLIRNIDPH-RTTANG-----DPTRIITHQVDIHGHNLQSIDP----- 60
Db 637 DGKKRIITRDFDQRETQEQYNDQDLVKIVQPGNGGIIRFAYNKQGNVIRKIDPEGSITWK 696
Qy 61 RLY-EAKQTNNITKP-NFLWQYDITGNPLCTESIDAG---RTVTINDIEGRPLLTVTATG 115

```

RESULT 11
 US-09-252-991A-23640
 ; Sequence 23640, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 107196.136
 ; CURRENT APPLICATION NUMBER: US/09/252.991A
 ; CURRENT FILING DATE: 1999-02-18
 ; PRIOR APPLICATION NUMBER: US 60/074.788
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 60/094.190
 ; PRIOR FILING DATE: 1998-07-27
 ; NUMBER OF SEQ ID NOS: 33142
 ; SEQ ID NO 23640
 ; LENGTH: 974
 ; TYPE: PRT
 ; ORGANISM: Pseudomonas aeruginosa
 ; US-09-252-991A-23640

Db 1215 GRITQDPGLKGGWNFYQYPL-NPVNTNDPIGL 1247

RESULT 13
US-09-543-681A-4476
; Sequence 4476, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 4476
; LENGTH: 1665
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-4476

Query Match 5.2%; Score 261.5; DB 2; Length 1665;
Best Local Similarity 21.1%; Pred. No. 2.6e-10;
Matches 184; Conservative 124; Mismatches 348; Indels 215; Gaps 42;

Qy 31 HRTTANGDPDTRITHQYDIHGLHQSIDP---RLYAKOTNNVTK---PN-FLWQYDL 82
Db 816 HFTSAFGDE---WQLSVDENGRLIVTDPOQRQVYBSQHGELLKAITPAGQWQHY 871
Qy 83 -----TGNNPL---CTE---SIDAGRTVTLNDIEGRPLLTATVATGVIOTROYETSSLPGR 130
Db 872 NPAHQLIKNTNPNYQHSSTVHSDELGLLHYTD-----ALNHTTRYQYS----- 914
Qy 131 LLSVAEQTPPEKTSRITERLIWAGNTEAKDHNLAGQCVRHYDTAGVTRLESLSLGTVL 190
Db 915 -----TEHASTNGSLSKILLPDGVEQIDYDSERRVIAVTDGEGKTRYRYGPPFDVLL 967
Qy 191 SQSSOLLIDTOBANWGTNETWQNMADDIVTTLSTPDATGALLTQTDAGNIQRLAYD 250
Db 968 ---AMIRPDGSEIRFEYDLSLTKKVNANGEVLYERDKAGIIRVDFTGRCIYRD 1024
Qy 251 VAG-----QLANGSWLTL-----KQTEQVIAKSLTYS-AAGQKLREBHG 288
Db 1025 RLGRRIATRYPDNHELRYNESGLVQSEWFEDEQSSRCLSTAQYVNAQQLIKATN 1084
Qy 289 NDVITEYSYEP-----TORLIGIKTRRPSDKVL-----QDLRYEYDPVGNVISIR 335
Db 1085 PDSVVEFEYDDQGRLCSEIRINEQEIHVQWNEADNTLALTRFGERELHYAFGALGELSLQ 1144
Qy 336 -NDAEATRFWN---QKWPENT---YTVDSLYQLISATGREMANIGQQSHQPPSPALPSD 389
Db 1145 VNQHAPLOFSYNAVQCEYLRRSRAGFVNSHYTATGLLAHGRAGTGTFQSIQAHFQQ 1204
Qy 390 NNTYNTYRTYTRDGGMLTKIQHS-----SPATON----- 420
Db 1205 PPFCTDVHSYQDRAYNVVGIEDDRWQTRVHYNANDQITETQYSPQWQDEKFOYDN 1264
Qy 421 ---NYTNTIT-----VSNRSNRAVLSTLTEDPA-QVDALFDAGGHQNTL--ISGO----- 464
Db 1265 NLNITEHLTTPSSMVPDQAQMLQLFQQQAGRVTRRYTAKGYQDYHYDVNGRLAKKI 1324
Qy 465 -----NLNNTTRELQVTLVKRDKGANDREW-VRYSGDGRMLKINEQQAASN 512
Db 1325 VHTGRFRERWYLNNTQNLFACTPPKDC-----WHYTFDAGRLSK-TKTVUSD 1376
Qy 513 NAQTORVYLPNLERLTQ-----NSTATTEDLOVITVGEAGRA-QVRVLHWSGKPRDI 566
Db 1377 LAHIDLPL--FPQIKPKITWYLVGSDQLIBETPIYADGTLANAQVQWLY---QPGEI 1430
Qy 567 DNNQLRYSDNLIQSSQLELSEGIISIEEYYPYGGTALWAAR----- 610

Db 1431 -TPTARYOQGL---HYVVDHQG---TPREIFSEGGQASWAGRLNTWQMQFWRYRDK 1483

Qy 611 -----NOTEASYKTIYSGK-ERDATGLYYGYRYQYQWIGRWLSSDPAGTIDGLNLVYM 664
Db 1484 AENDPNYTECPF---RFAGQYDEBESGLYNNRFRYYDRETQYLSFDPGLGLGGLNPGY 1540
Qy 665 VRNPNVTLDDPGLMPTTAERIAALKXN---KYTDSAPSANATN-VAINIRPPVA--- 716
Db 1541 V-HCPTGWVDPFGLAGDCDELLVLATNPRSKTVQEMAEDLANRINRNTVTFTSTPSMGH 1599

Qy 717 ---PKPSLPKASTSSQPTTH-PIGAANIKP 742
Db 1600 IDLRGRSHFDKVTQADIPTPHVQSPINAP 1630

RESULT 14
US-09-251-645-6
; Sequence 6, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 6
; LENGTH: 1584
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-6

Query Match 4.9%; Score 245; DB 2; Length 1584;
Best Local Similarity 19.3%; Pred. No. 4.2e-09;
Matches 202; Conservative 126; Mismatches 307; Indels 414; Gaps 44;

Qy 14 TVSYDNRGLIRID---FHRTTANGDPDTRITHQYDIHGLH-NQSIDPRLYEAKQTN 69
Db 552 TVTGFDAHMSKNTVSIYTHQLRKVDNVHVIDQSYDLGLGRITGQIIDP-----GTA 605
Qy 70 NTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTATVATGVIOTROYETSSLP 129
Db 606 REIKRNVYQY-----PGGDE-----NDF-WPVMIEVDSQGVRRKTHYDGM--G 646
Qy 130 RLLSVAEQTPPEK---TSRI---TERLIWAGNTEAKDHNLAGQCVRHYDTAGVTRLE--- 181
Db 647 RICSIEODDGMAGCTSGIYQGYRKVLA-----ROYDLVQLSKESLN 690
Qy 182 ---SLSTCTVLSQSSQLLIDTQEANWTDNETWQNMADD----- 220
Db 691 DMLNLSANPLVRLATPLVTKYKYG-----MGNLYSTEYSDGRIEIHDPITRTIT 745
Qy 221 -----YTTLS-TFDTAGALLTQTDAGNTOQLAY 249
Db 746 QGVKGLGLMLNIQNNFEQASIKAVYPDGTIYSTRYDGFRTVTETDAEGHATQIGY 805
Qy 250 DV-----AGOLNGSWL-----TLKQOT 266
Db 806 DVFRIVKTLPDGTILESAYASPSHEELISALNVNGTQGLNLYVDGLGRVLSIDTVGRK 865
Qy 267 EQVI-----IKSLT-----YSAAGQKLRE 286
Db 866 TEYLYGPGQDKPIQSIPTSHNKQNMVLYLGSVMSKFTTGTQDQNFYHSHKGTLLSAS 925
Qy 287 HGNDVITEYSYEPE--TORLIGIKTRRP-SDTKVLQDUR-----YEYDPV 328

Db 926 EGVSQ-TNYSYFPGVLOREFLRDNKPISEGEVLYTMSGLIQRKDSFGHNHYSDAQ 984
Qy 329 GNVISIRNDABATRFMHNQKMPENTYDYSLYQLISATGREMANIGO-----376
Db 985 GRLVKTEQDAQYA-----TPEYDNVGRLLITTTTKDTTSLSLQVTKIEYDADR 1032
Qy 377 -----QSHQPPSPAL 386
Db 1033 EIKRSLISDFSQVITLSYTKNNOISORITSIDGVMMKERYQYDNNORLSQYQCEGEQS 1092
Qy 387 PSDNNTYNTYRTYTDYRGGLTKIOHSSPATONNTT-NITVS---NRSNRAVLSTLITE 442
Db 1093 PIDHTRVLNQOIYHYQWGNIKEL-----DNYRDGKETVDYHFSQADPTQLIRITS 1145
Qy 443 DPAQVDALFDAGGHONTLISQNLNMTNRGLQVTLVKRDKGANDREWYRSGDGRM 502
Db 1146 DKQIELSYDANGNLTRDEKGTLLIYDQNNRLVQV---KDLGN-----1186
Qy 503 LKINEQASNAQORVTYLPNLELRLTONSTATTEDLOVITVGEAGRAQVRVLHWESGK 562
Db 1187 LVCSYQVDALNKLIT-----AQVLANGTVNRQ-----HYASGK 1218
Qy 563 PEDIDNNQLRYSYDNLIGSSQLELDSQIIESEBEPYIG---GTALWAARNOTE---AS 616
Db 1219 ---VTNQLQDEALITSLSSDKORIGHOSAKNGQSVYYQYIGIDHNSTVIASQNELMALIS 1275
Qy 617 YKTIRY-----SKERD-ATGLYYY--GYRYQYPTWGRWLSSD---PAGTIDGLN 660
Db 1276 YTPYGFRLSISLPLNGAQVDPVTGWYFLNGYRVFENPVLMRPHSPDSWSPFGR-GGIN 1334
Qy 661 LYRMVRNPNVTLDPDGLMPT-----TAERIAALKKKNVTDAPSANATVA 708
Db 1335 PYTYCGDPPINRIDLNGHLSAGGILVGAIGIIVGLSVGAGAAISAGLIAAGGALGA 1394
Qy 709 INIRPPVAPKPSLKASTSSOPT-----HPIGAANI-----KPTTSG-----SSIV 750
Db 1395 I-----ASTSALAVTATVIGLAADSIGIASAALSEKDPKTSGLNWSIAGL 1440
Qy 751 APLSPVGNKTSISLPESAQSSSSTTS 779
Db 1441 GVLSFGISAITFTSSLVKSARSQSAVSA 1469

RESULT 15
US-09-252-991A-31032
; Sequence 31032, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 31032
; LENGTH: 1128
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31032

Query Match 3.9%; Score 193.5; DB 2; Length 1128;
Best Local Similarity 18.3%; Pred No. 1.8e-05;
Matches 181; Conservative 110; Mismatches 301; Indels 395; Gaps 41;
Qy 4 IDPK-----LYKQKPT-----VSVYDNRGLIIRNIDF---HRTTANGDPPDTRITRHOYDI 50
Db 257 IDPQSGLDMLVRYQFSDSGDLIAVRDLRGQVRVFAWREHMLVAHGEPGLGVREYWDV 316

Search completed: February 16, 2006, 21:46:40
Job time : 27.6251 secs

Qy 51 HGHLNQSIDPRLYEAKQTNNTIKENFLWOYDLTGNPLCTESIDAGRTVTTLNDISGRPLLT 110
Db 317 HAP-----HGRVVKQ 326
Qy 111 VTATGVITQTRQY-----ETSSLPGRLSVABQPEEKTSRITERLIIWAGNTEAKOHNLA 165
Db 327 IEAGGLTRTRYLRDTEVSDSLGRV--ERYEPAGEGQRRWTALVRADGSRSEFDYDLF 384
Qy 166 GQCVRHVDTAGVTRLESLSUTGVVLSOSSQ-----LLDITQBA---NWT-- 206
Db 385 GRLVAMRDPLGRETRRRRRGQGRMLEESPGRYKRVRDEETGLLVELEDAMQRRWTFE 444
Qy 207 -----GNETVWQNMLADDITVTLIS-- 226
Db 445 RDERGNATTVRGPGAGSTRYAYEDRPLDRPTRIVDPGRGERRLEWRNFGLLAULTDCSGQ 504
Qy 227 ---TFDATGALLTQTDAGNIQRLAYDAGQL-----NGSWITLKGOTEQVIKSLTY 276
Db 505 VMRYDYNEGRLVASSDPLGQLTERRYDPLGQLIGLELADGSAL-----SVEY 552
Qy 277 SAAGQKUR--BEHGNVITEYSYEPETORLIGIKTRRPSDTKVLODLRYEYDVPQNVISI 334
Db 553 DALGROTRIADAEGHATLFSWGH-----GDLARVSDAGG-GELSYLHDEAGRLVAL 603
Qy 335 RNDAEATRFWNNQKMPENTYTDLSYQLISATGREMANIGQOSHOPSPALPDNNNTYT 394
Db 604 TNENGV-----QAQFRYDLDLRLVEETGDFGR---RQRYRNAADELAREDDAD 649
Qy 395 NYTRTYTYDRGGLNLTQIHSPPATONNTTNTVSNRSRAVLSTLTEDPAQVDALFDAG 454
Db 650 GRETTYAYDRGLASIR--VPATE----- 672
Qy 455 GHQNTLISQNLNMTNRGELQVTLVKRDKGANDREWYRSGDGRMLK-----IN 506
Db 673 -HAPALV--BRYRWLADGLASA-----GGADCEVRYTYDEVGNLRLESQVHADGWVY 722
Qy 507 EQQASNNA---OTOR-----VTYLPN-----LEL-----R 528
Db 723 SVEHSHDALGVQRTSRYGDAPPVAVLTYGPHLLHGLVGAVELAFERDALHREVRDARR 782
Qy 529 LTQNSTATTEDLVITVGEAGRAQVRV---LHWESGKPED-----IDNNQ---LRYSY 575
Db 783 DGQDDALFTQERQHAPLQRLQSRRLAGGDFWQGVYDGLGQLVGDIDNNQVPSVRYEY 842
Qy 576 D---NLIGS-----SOLELDSG-----QIIESEYYPYG----- 602
Db 843 DLGRLLASRRAGAAASTYRYDAAGNRLEGVGEYAREDAQAFAENELYSRSGFSRSEVRA 902
Qy 603 ---GTALWAA-----RNQTEASYKTYRYSKGERDATGLYYYG----- 636
Db 903 NQAGEGPAPWAGNVERIAGNRYRFDALGNLIVERIGADGELRLAYDGAQRLVHLTRDYA 962
Qy 637 -----YRYQ-----PWIG-----RWLSSDPAGTI 656
Db 963 DGTRELEARYDALSRRRIAKVLRDGVQVRFQWMDGROCAEAFARELRITTVHEPGGFV 1022
Qy 657 DGLNLYRMVRNPNVTLID-----PDGLMPTIARIALKKNKV-----TDSAPSPANATN 706
Db 1023 PLLLEQACEFPDPPPELLQLRQFAAEG-QFLPAQCVCPALGEARIAFTHTDLHGTPLQLSD 1081
Qy 707 VAINIRPPVAP---KPSLPKASTSSQP 730
Db 1082 ERGQLRWQGVDDWRVAPERQGAQP 1108

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:22:44 ; Search time 146.155 Seconds
(without alignments)
4431.220 Million cell updates/sec

Title: US-10-754-115-45
Perfect score: 7901
Sequence: 1 MQNSQDFSLTSLPLKGGGA.....WFTVNEDENTAAEYKVKVM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq 21:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7901	100.0	1474	8	ADP18622 Photorhab
2	7901	100.0	1474	8	Adt21525 Photorhab
3	7285.5	92.2	1476	6	Abm70222 Photorhab
4	6065.5	76.8	1477	6	Abm70230 Photorhab
5	6042	76.5	1481	2	Aay33730 Photorhab
6	6031.5	76.3	1476	5	Abg32652 P. lumine
7	6025.5	76.3	1476	8	Adt21576 Photorhab
8	4627.5	58.6	1486	6	Abm70526 Photorhab
9	4595.5	58.2	1485	8	Adt21578 Photorhab
10	4581.5	58.0	1485	2	Aaw56546 Toxin tca
11	4580.5	58.0	1485	2	Aaw17887 Photorhab
12	4261	53.9	1506	8	Adt21529 Xenorhabd
13	4261	53.9	1506	9	Aeb47812 Native Xp
14	3971	50.3	1428	8	Aay97695 SepB prot
15	3971	50.3	1428	8	Adt21540 Serratia
16	3834	48.5	1493	8	Adr20369 Recombina
17	3834	48.5	1493	8	Adt21498 Xenorhabd
18	3106.5	39.3	1444	8	Adm61347 Paenibaci
19	3106.5	39.3	1444	8	Adt21520 Paenibaci
20	1744.5	22.1	662	3	Aay95685 Cosmid cH
21	1262	16.0	697	8	Adm61375 Paenibaci
22	839	10.6	591	2	Aar23006 Protein t
23	199	2.5	2060	5	Aae20967 Staphyloc
24	172.5	2.2	1404	7	ADC00960 Enterohae

25	170	2.2	593	6	AAO19846	Aao19846 P aerugin
26	169.5	2.1	14130	9	AEA07438	Aea07438 M. ulcera
27	162.5	2.1	16990	9	AEA07436	Aea07436 M. ulcera
28	161	2.0	1863	7	ADD46989	Add46989 Rat prote
29	161	2.0	1863	9	ADX28456	Adx28456 Novel cel
30	160.5	2.0	1385	6	ABU40318	Abu40318 Protein e
31	160.5	2.0	1426	3	AAAB15983	Aab15983 E. coli p
32	160.5	2.0	1426	6	ABU14693	Abu14693 Protein e
33	159.5	2.0	1329	4	ABG28402	Abg28402 Novel hum
34	159.5	2.0	1329	4	ABG25780	Abg25780 Novel hum
35	159.5	2.0	1329	7	ADC33172	Adc33172 Human nov
36	155.5	2.0	658	7	ABO76164	AbO76164 Pseudomon
37	155.5	2.0	1400	7	ADC01365	AdC01365 Enterohae
38	155.5	2.0	2315	7	ADF05149	AdF05149 Bacterial
39	154	1.9	4630	3	AAAY77177	Aay777177 S. venezu
40	154	1.9	5215	8	ADL91930	Adl91930 Streptomy
41	153.5	1.9	1028	7	ADF06896	Adf06896 Bacterial
42	152.5	1.9	1669	8	ADU07870	Adu07870 Amino aci
43	152	1.9	4630	2	AAW19629	Aaw19629 Streptomy
44	151.5	1.9	1904	5	ABB57100	Abb57100 Mouse isc
45	151.5	1.9	1904	9	ADX26385	Adx26385 Novel cel

ALIGNMENTS

RESULT 1

ADP18622
ID ADP18622 standard; protein; 1474 AA.

XX AC ADP18622;

XX DT 12-AUG-2004 (first entry)

XX DE Photorhabdus luminescens W-14 tcdB2 protein SEQ ID NO:10.

XX KW tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;

XX KW orally active insect toxin; insect toxin; tcdB2.

XX OS Photorhabdus luminescens.

XX PN WO2000404217-A2.

XX PD 27-MAY-2004.

XX PF 12-NOV-2003; 2003WO-IB005553.

XX PR 12-NOV-2002; 2002US-0425672P.

XX PA (UYBA-) UNIV BATH.

XX PI Ffrench-Constant RH, Waterfield NR;

XX DR WPI; 2004-411735/38.

XX DR N-PSDB; ADP18621.

XX PT New isolated Photorhabdus luminescens nucleic acids, useful for expressing orally active insect toxin or for generating transgenic plants with enhanced resistance to insects.

XX PS Claim 1; SEQ ID NO 10; 118pp; English.

XX CC The present invention describes DNA sequences from the tcd genomic region of Photorhabdus luminescens W-14. Also described: (1) a transgenic monocot or dicot cell having a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant with a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a progeny of seed; (5) a method of producing Toxin A of Photorhabdus luminescens W-14 in a heterologous host; and (6) a method of producing an orally active insect toxin. The nucleotide sequences are useful for heterologous expression of orally active insect toxin. They can also be used for generating transgenic plants with enhanced resistance to

CC insects. The present sequence represents Phototribus luminescens W-14
 CC tcdB2, which is used in the exemplification of the present invention.
 XX
 SQ Sequence 1474 AA;

Query Match 100.0%; Score 7901; DB 8; Length 1474;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1474; Conserved 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSGDFSIITSLPKGGGALITGGEALPTGPDGMAALSPLPISAGRGYAPAFITLNN 60
 DB 1 MNSGDFSIITSLPKGGGALITGGEALPTGPDGMAALSPLPISAGRGYAPAFITLNN 60

QY 61 SGAGNSPGLGWCNMTIRRTIRRTIRRTIRRTIRRTIRRTIRRTIRRTIRRTIRRT 120
 DB 61 SGAGNSPGLGWCNMTIRRTIRRTIRRTIRRTIRRTIRRTIRRTIRRTIRRT 120

QY 121 LGATFTVTGYSRLESFSLRLEYWQPKTTGTDWLIYSPDGQVHLLGKSPQARISNPSQ 180
 DB 121 LGATFTVTGYSRLESFSLRLEYWQPKTTGTDWLIYSPDGQVHLLGKSPQARISNPSQ 180

QY 181 TTQTAQWLLASVSRSRGQIYYVRAEDDTGCEADEITHHLQATAQRYLHIVYGNRTAS 240
 DB 181 TTQTAQWLLASVSRSRGQIYYVRAEDDTGCEADEITHHLQATAQRYLHIVYGNRTAS 240

QY 241 ETLPGLDGSAPSQADWLEFVDFYDGERSNLKTTPAFSTTGSWLCRODRFSRYGFEIR 300
 DB 241 ETLPGLDGSAPSQADWLEFVDFYDGERSNLKTTPAFSTTGSWLCRODRFSRYGFEIR 300

QY 301 TRRLCRQVLYHHLQALDSKITEHNGPTLVSRLLIINYDESAIATLVFVRVHGQDGNV 360
 DB 301 TRRLCRQVLYHHLQALDSKITEHNGPTLVSRLLIINYDESAIATLVFVRVHGQDGNV 360

QY 361 VTLPLELAYQDFSPRHHAHQPMQDVLANFNAIQRWQLVDLKGELPGLLYQDKGAWWYR 420
 DB 361 VTLPLELAYQDFSPRHHAHQPMQDVLANFNAIQRWQLVDLKGELPGLLYQDKGAWWYR 420

QY 421 SAQRLEIGSDAVTWKQOPSLVPSLOSASLVNDINGDGLDQVITGPGLRGHSORPD 480
 DB 421 SAQRLEIGSDAVTWKQOPSLVPSLOSASLVNDINGDGLDQVITGPGLRGHSORPD 480

QY 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ 540
 DB 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVVQ 540

QY 541 SGDTITLPPGADPRKLVAFSDVLSGQAHLEVSATKVTCPNLRGRFGQPIITLPGFSQ 600
 DB 541 SGDTITLPPGADPRKLVAFSDVLSGQAHLEVSATKVTCPNLRGRFGQPIITLPGFSQ 600

QY 601 PATFENPAQVYLADLDGSGPTDLIVVHTNRDLIFLKSNGFGFAPVTLRFPPEGLRFDHTC 660
 DB 601 PATFENPAQVYLADLDGSGPTDLIVVHTNRDLIFLKSNGFGFAPVTLRFPPEGLRFDHTC 660

QY 661 QLQWADVOGLGVASLILSVPHMSPHHWCDDLTKMKPMLLNEMNNMNGVHHTLYRSSSQF 720
 DB 661 QLQWADVOGLGVASLILSVPHMSPHHWCDDLTKMKPMLLNEMNNMNGVHHTLYRSSSQF 720

QY 721 WLDEKAAALTTGQTPVCVLPPIHTLMOTETEDISGNKLVTTLRYARGAWDGRERFRG 780
 DB 721 WLDEKAAALTTGQTPVCVLPPIHTLMOTETEDISGNKLVTTLRYARGAWDGRERFRG 780

QY 781 FGYYEOTDSHOLACGNAPERTPPALTKWNYATGLPVIDNALSLEYWRDDQAFAGSPRFT 840
 DB 781 FGYYEOTDSHOLACGNAPERTPPALTKWNYATGLPVIDNALSLEYWRDDQAFAGSPRFT 840

QY 841 TWQNDKQVPLTPEDDNSRYWFNRLKGLLSELYGLDSTNKHVPYTVTFEFSQVRLQ 900
 DB 841 TWQNDKQVPLTPEDDNSRYWFNRLKGLLSELYGLDSTNKHVPYTVTFEFSQVRLQ 900

QY 901 HTDSRYPVLSWVESRNYHRTIASDPQCSQNTILSSDRFGQPLKQLSVQYPRQQPAI 960
 DB 901 HTDSRYPVLSWVESRNYHRTIASDPQCSQNTILSSDRFGQPLKQLSVQYPRQQPAI 960

QY 961 NLYPDTLPDKLLANSYDDQQRQLRLTYQQSSWHLLTNNTVRVVLGLPDSRSDIFTYGAEN 1020
 DB 961 NLYPDTLPDKLLANSYDDQQRQLRLTYQQSSWHLLTNNTVRVVLGLPDSRSDIFTYGAEN 1020

QY 1021 VPAGGLNLELISDKNSLIADDKPREYLGQOKTAYTDGONTTPTPTPQALIAFTETTVF 1080
 DB 1021 VPAGGLNLELISDKNSLIADDKPREYLGQOKTAYTDGONTTPTPTPQALIAFTETTVF 1080

QY 1081 NOSTLSAPNGSIPSDKLSLTLLEQAGYQOQTNVLPRTGEDKVVVAHGVTDYGTAAQFWRP 1140
 DB 1081 NOSTLSAPNGSIPSDKLSLTLLEQAGYQOQTNVLPRTGEDKVVVAHGVTDYGTAAQFWRP 1140

QY 1141 KQSQNTQLTKITLIMDANYCVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDNQHLITL 1200
 DB 1141 KQSQNTQLTKITLIMDANYCVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDNQHLITL 1200

QY 1201 DALGRPTILRFPWGTENGKMTGYSPEKASPPSDVNAAIELKKPLPVAQCVVAPESWM 1260
 DB 1201 DALGRPTILRFPWGTENGKMTGYSPEKASPPSDVNAAIELKKPLPVAQCVVAPESWM 1260

QY 1261 PVLISOKTFNRLAEODWOKLYNARIITEDGRICTLAYRRWVQSOKAIPOLISLLNNGPRLP 1320
 DB 1261 PVLISOKTFNRLAEODWOKLYNARIITEDGRICTLAYRRWVQSOKAIPOLISLLNNGPRLP 1320

QY 1321 PHSLLTTRDYDHDPEQIIRQVVVSDGFGRLQLQAAARHEAGMARQNRNEDGSLIINVQHT 1380
 DB 1321 PHSLLTTRDYDHDPEQIIRQVVVSDGFGRLQLQAAARHEAGMARQNRNEDGSLIINVQHT 1380

QY 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFNLWRYVNSDSARQEKAYADTHVYDPIGREIKV 1440
 DB 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFNLWRYVNSDSARQEKAYADTHVYDPIGREIKV 1440

QY 1441 ITAKGWFRRLTFTPWFTVNEDENDTAAEVKKVKM 1474
 DB 1441 ITAKGWFRRLTFTPWFTVNEDENDTAAEVKKVKM 1474

RESULT 2
 ADR21525
 ID ADR21525 standard; protein; 1474 AA.
 XX ADR21525;
 AC
 XX
 DT 04-NOV-2004 (first entry)
 XX
 DE Phototribus tcdB2 toxin SEQ ID NO:45.
 XX
 KW toxin; insect; insecticidal; transgenic; pest control.
 XX
 OS Phototribus luminescens.
 XX
 FN WO2004067727-A2.
 XX
 PD 12-AUG-2004.
 XX
 PF 07-JAN-2004; 2004WO-US0000394.
 XX
 PR 21-JAN-2003; 2003US-0441723P.
 XX
 PA (DOWC) DOW AGROSCIENCES LLC.
 XX
 PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
 PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
 XX
 DR WPI; 2004-580999/56.
 DR N-ESDB; ADR21524.
 XX
 PT Controlling or inhibiting an insect, useful for pest control, comprises
 PT contacting the insect with effective amounts of a Protein A, a Protein B,
 PT and a Protein C.
 XX
 PS Claim 1; SEQ ID NO 45; 368pp; English.
 XX

CC The invention relates to a novel method for controlling or inhibiting an
 CC insect comprising contacting the insect with effective amounts of a
 CC protein A, a protein B, and a protein C. Each of the proteins A, B, and C
 CC is encoded by a naturally occurring gene or has an amino acid sequence
 CC that differs from the product encoded by a naturally occurring gene only
 CC by truncation or by conservative amino acid changes. Protein A is a 230-
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic
 CC species, has stand alone insecticidal activity, and has an amino acid
 CC sequence at least 40% identical to a sequence selected from XptA2wi,
 CC XptA2wi, TcdA, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin
 CC complex potentiator having an amino acid sequence at least 40% identical
 CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wi, XptB1xb,
 CC PpC1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
 CC having an amino acid sequence at least 35% identical to a sequence
 CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptC1xb, PpC1
 CC (orf 6 long), PpC1 (orf 6 short), and SepC. Also claimed is a transgenic
 CC plant or plant cell that produces a protein A, a protein B, and a protein
 CC C. The method is useful for pest control. The present sequence represents
 CC Photorhabdus luminescens TcdB2 toxin.

XX SQ Sequence 1474 AA;

Query Match 100.0%; Score 7901; DB 8; Length 1474;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONSQDFISITSLPKGGGAIATGMEALTPGPDGMAALSPLPISAGRGYAPAFITLNYN 60
 DB 1 MONSQDFISITSLPKGGGAIATGMEALTPGPDGMAALSPLPISAGRGYAPAFITLNYN 60
 QY 61 SGAGNSPGLGWDNCVMTIRRTTHRGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120
 DB 61 SGAGNSPGLGWDNCVMTIRRTTHRGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120
 QY 121 LGATFTVTVGSRLESFHSRLYEQPKTKTDFWLIYSPDQVHLLGKSPQARTSNPSQ 180
 DB 121 LGATFTVTVGSRLESFHSRLYEQPKTKTDFWLIYSPDQVHLLGKSPQARTSNPSQ 180
 QY 181 TTQTAQWLLASVSRGQIYYQRAEDDTGCEADEITHLQATAQRYLHVYGNRTAS 240
 DB 181 TTQTAQWLLASVSRGQIYYQRAEDDTGCEADEITHLQATAQRYLHVYGNRTAS 240
 QY 241 ETLPGDGSAPQADWLFYLPVDFYGERSNLKTTPPAFTTGSWLCRQDRFSRYGPEIR 300
 DB 241 ETLPGDGSAPQADWLFYLPVDFYGERSNLKTTPPAFTTGSWLCRQDRFSRYGPEIR 300
 QY 301 TRRLCROVLYMHHLOALDSKITEHNGPTLVSRLLIYNDSEALASTLVFVRVHGEGQGV 360
 DB 301 TRRLCROVLYMHHLOALDSKITEHNGPTLVSRLLIYNDSEALASTLVFVRVHGEGQGV 360
 QY 361 VTLPPLELAYQDFSPRHHAHQPMQDVLANFNAIQRWQLVLDKGEGLPGLLYQDKGAWMYR 420
 DB 361 VTLPPLELAYQDFSPRHHAHQPMQDVLANFNAIQRWQLVLDKGEGLPGLLYQDKGAWMYR 420
 QY 421 SAORLGEIGSDAVTWKMQPLSVTPSLOSASLVNDINGDGLDWTITGPGLRGTHSQRPD 480
 DB 421 SAORLGEIGSDAVTWKMQPLSVTPSLOSASLVNDINGDGLDWTITGPGLRGTHSQRPD 480
 QY 481 GSWTRFTPLNALPVETHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKGVVQ 540
 DB 481 GSWTRFTPLNALPVETHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKGVVQ 540
 QY 541 SGDTLTPVPGADPRKLVAFSDVLSGQAHLEVSATKVTCPNLRGRFGQPIITLPGFSQ 600
 DB 541 SGDTLTPVPGADPRKLVAFSDVLSGQAHLEVSATKVTCPNLRGRFGQPIITLPGFSQ 600
 QY 601 PATBNPAQVYLAIDLGGSPDILYVHTNRDLIFLNSGNGFAEPVTLRPFEGRLFDHTC 660
 DB 601 PATBNPAQVYLAIDLGGSPDILYVHTNRDLIFLNSGNGFAEPVTLRPFEGRLFDHTC 660
 QY 661 QLOWADVQGLGVASLILSVPHMSPHHWCDCITNNKPKWLLNEMNNMNVHHTLRYRSSQF 720
 DB 661 QLOWADVQGLGVASLILSVPHMSPHHWCDCITNNKPKWLLNEMNNMNVHHTLRYRSSQF 720

QY 721 WLDEKAAALTTGQTPVCYLPPIHTLMQTEDEBISGNKLVTTILRYARGAWDGRERBRFG 780
 DB 721 WLDEKAAALTTGQTPVCYLPPIHTLMQTEDEBISGNKLVTTILRYARGAWDGRERBRFG 780
 QY 781 FGYYEQTDSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEVWRDDQAFAGSPRPT 840
 DB 781 FGYYEQTDSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEVWRDDQAFAGSPRPT 840
 QY 841 TWQNDKDVLPTEPDNSRYWENRALKGOLLRASELYGLDDSTNKGVPPYTVTFRSQRRLQ 900
 DB 841 TWQNDKDVLPTEPDNSRYWENRALKGOLLRASELYGLDDSTNKGVPPYTVTFRSQRRLQ 900
 QY 901 HTDSRYPLVMSVSVESRYHYERIASDPQCSQNTLSSDRFGQPLKQLSVQPRRQPAI 960
 DB 901 HTDSRYPLVMSVSVESRYHYERIASDPQCSQNTLSSDRFGQPLKQLSVQPRRQPAI 960
 QY 961 NLYPDTLPDKLLANSYDDQORQLRTYQOOSWHLTNNTVRLVGLPDSSTRSDIFTYGAEN 1020
 DB 961 NLYPDTLPDKLLANSYDDQORQLRTYQOOSWHLTNNTVRLVGLPDSSTRSDIFTYGAEN 1020
 QY 1021 VPAGLNLLELSDKNSLIADDPREYLGQOQKATYTDGQNTTPTQTPTRQALIAFTETTVF 1080
 DB 1021 VPAGLNLLELSDKNSLIADDPREYLGQOQKATYTDGQNTTPTQTPTRQALIAFTETTVF 1080
 QY 1081 NQSTLSAFNGSIIPSDKLSSTLLEQAGYQOQNTYLPRTGEDKVVVAHHGYTDYGTAAQFWRP 1140
 DB 1081 NQSTLSAFNGSIIPSDKLSSTLLEQAGYQOQNTYLPRTGEDKVVVAHHGYTDYGTAAQFWRP 1140
 QY 1141 OKQSNTOITGKILTIWDANVCVVQTRDAAGLTTSKYDWRFLTPVQLTDINDNOHLITL 1200
 DB 1141 OKQSNTOITGKILTIWDANVCVVQTRDAAGLTTSKYDWRFLTPVQLTDINDNOHLITL 1200
 QY 1201 DALGRPTILRFWGTENGKMTGYSSEKASFPSPSDVNAIIEKKPLPVAQCVYAPESWM 1260
 DB 1201 DALGRPTILRFWGTENGKMTGYSSEKASFPSPSDVNAIIEKKPLPVAQCVYAPESWM 1260
 QY 1261 PVLQKTFNRLAEQDWQKLYNARIITEDGRCTTAYRRVWOSQKAIPLIISLNNNGPRLP 1320
 DB 1261 PVLQKTFNRLAEQDWQKLYNARIITEDGRCTTAYRRVWOSQKAIPLIISLNNNGPRLP 1320
 QY 1321 PHSITLTTDRVDHDPQOIRQOVVPSDGFGRLLQAAARHEAGMARQDNEDGSLIINVQHT 1380
 DB 1321 PHSITLTTDRVDHDPQOIRQOVVPSDGFGRLLQAAARHEAGMARQDNEDGSLIINVQHT 1380
 QY 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHVYDPIGREIKV 1440
 DB 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHVYDPIGREIKV 1440
 QY 1441 ITAGWGFRTLTFTPWFTVNEDENTAAEVKVKVM 1474
 DB 1441 ITAGWGFRTLTFTPWFTVNEDENTAAEVKVKVM 1474
 RESULT 3
 ABM70222
 ID ABM70222 standard; protein; 1476 AA.
 XX AC ABM70222;
 XX DT 20-NOV-2003 (first entry)
 XX DE Photorhabdus luminescens protein sequence #3319.
 XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 XX KW detection; food; gene expression; plant; animal; microorganism; toxin;
 XX KW antibiotic; biopesticide; virulence factor; disease model; plague;
 XX OS whoping cough.
 XX OS Photorhabdus luminescens.
 XX PN WO200294867-A2.
 XX XX

[illegible]

RESULT 4

ABM70230
ID ABM70230 standard; protein; 1477 AA.

XX AC ABM70230;

XX XX 20-NOV-2003 (first entry)

XX DE Photorhabdus luminescens protein sequence #3327.

XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX KW detection; food; gene expression; plant; animal; microorganism; toxin;
XX KW antibiotic; biopesticide; virulence factor; disease model; plague;
XX KW whooping cough.

XX OS Photorhabdus luminescens.

XX PN WO200294867-A2.

XX XX 28-NOV-2002.

XX XX 07-FEB-2002; 2002WO-IB003040.

XX XX 07-FEB-2001; 2001FR-00001659.

XX XX (INSP) INST PASTEUR.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX XX Duchaud E, Taurit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX PI Buchrieser C;

XX XX WPI; 2003-148459/14.

XX XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX PS Claim 2; SEQ ID NO 3327; 1205pp; French.

XX CC The invention relates to the isolation of genes and their encoded
XX CC proteins from Photorhabdus luminescens. The isolated sequences are
XX CC sources of probes and primers for detecting the genome of P. luminescens
XX CC and related species; to study polymorphisms; for gene analysis and for
XX CC detection/amplification of the genes. Antibodies (Ab) raised against the
XX CC polypeptides encoded by the genes are used for detection/identification
XX CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX CC carry a gene-containing vector are used to select compounds that
XX CC modulate, regulate, induce or inhibit expression of the genes in plants,
XX CC animals or microorganisms other than P. luminescens and are able to alter
XX CC response or sensitivity to toxins and antibiotics produced by P.
XX CC luminescens. Cells transformed to express the genes are useful for
XX CC recombinant production of the proteins, particularly toxins and
XX CC antibacterials useful as insecticides, bactericides and fungicides. The
XX CC genes, proteins, vectors containing the genes and Ab are also useful
XX CC therapeutically (to treat microbial infection by bacteria or fungi that
XX CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX CC biopesticides. Other uses of the genes and the proteins are as virulence
XX CC factors and for identifying targets of human diseases for which P.
XX CC luminescens is a model (particularly plague and whooping cough). This
XX CC sequence represents one of the isolated P. luminescens proteins

XX SQ Sequence 1477 AA;

Query Match 76.8%; Score 6065.5; DB 6; Length 1477;

Best Local Similarity 76.3%; Pred. No. 0;

Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;

QY 1 MONSQDFISITSLPKGGGATGMEALTPTGPDGMAALSPLPISAGRGVAPFTLNYN 60

DB 1 MONSQDFISITSLPKGGGATGMEALTPTGPDGMAALSPLPISAGRGVSPSLTNS 60

QY 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114

Db 61 SGAGNSPFLGWCNVMVIRRRTHFGVPHYDETDTFLGPEGEVLVAIALNENGQADIRSES 120
QY 115 TLOGINLGATFTVTGYSRSLRSHFSLRLEYWQPKTKTDFWLIYSPDQVHLLGKSPQAR 174
Db 121 SLOGINLGEFTVTGYSRSLRSHFSLRLEYWQPKTKTDFWLIYSPDQVHLLGKSPQAR 180
QY 175 ISNPSQTTQTAQWILLEASVSRGQIYYQYRAEDDTGCEADEITHLQATAQRYLHYVY 234
Db 181 ISNPLNVSQTAQWILLEASVSRGQIYYQYRAEDTNCETDEFTAHPNATVQRYLQAVHY 240
QY 235 GNRFASETLPGLDGSAFSDADWLVFVDFYGERSNLTKTPAFSTTGSWLCRQDRFSRYE 294
Db 241 GNLTASEVFTPLNGDDPLKSGWLFCLVPDYGERKNSLSEIPFPAKASSLWLCRQDRFSRYE 300
QY 295 YGPIRTRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAITASTLVFVRRVGH 354
Db 301 YGFELTRRLCRQILMFHRLQTLGQAKGDDDEPALVSRLLIDYDENAVISTLVSVRRIGH 360
QY 355 EODGNVTVLPPELAYQDFSPRHHAHQPMQMDVLANFNAIQWOLVDLKGEGPLGLLYODK 414
Db 361 EDNNTVLSLPPLELAYQDFPEPEQKARWQSMQMDVLANFNAIQWOLVDLKGEGVPGVLYQDR 420
QY 415 GAWYRSQAQRLGEIGSDAVTWKMQPLSVIPLSQNASLVLDINGDQGLDWITGGLRGY 474
Db 421 NGWYRSQAQROAGBEMNAVWGRMQLPITPALQDNASLMDINGDQGLDWITGGLRGY 480
QY 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAK 534
Db 481 HSQHPDGSWTRFTPLDALPIEYSHPRAQLADLMGAGLSDLVLIGPKSVRLYANNRDGFTQ 540
QY 535 GKDVQSGDITLVPFGADPRKLVAFSDVILGSGQAHLEVEVSATKVTCPNLCRGRGQDIT 594
Db 541 GRDVQSGDITLPLPGADARKLVAFSVLGSGQAHLEVEVSATQVTCWPNLGHGREGQDIT 600
QY 595 LPPGFSQAPATENPAQVILADLDGSGPTDLIVVHTNRDLDFLNKSGNGFAEPVTLRFPBGL 654
Db 601 LPPGFSQADNPNDRVHLADLDGSGPADLIIVHTRLEIFKNESGNSFAKFTLRFPPGL 660
QY 655 RFDFTCOLQWADVOGLGVASLILSVPHMSPHHRCDLTNMKPWLNLNENNMNVHHTLRY 714
Db 661 RFDFTCOLQWADVOGLGVASLILSVPHMAPHHRCDLTNAPFWLLSEMMNMGAAHTLHY 720
QY 715 RSSQFQWDEKAAALATGCTQTPVCVLPPIHTLWQTEDEISGNKLVTLIYARGAWDGR 774
Db 721 RSSQFQWLDKAAALATGCTQTPVCVLPPEVHTLWQTEDEISGNKLVTLIYARGAWDGR 780
QY 775 EREPRGCGYVQTDHSHQLAQGNAPERTPPALTQWYATGLPVIDNALSTEYWR-DDQAFA 833
Db 781 EREPRGCGYVQTDHSHQLAQGNAPERTPPALTQWYATGLPVIDNALSTEYWR-DDQAFA 840
QY 834 GFSPRFTTWQDNKDVPLTPEDDNRYSRYWFRNALKQLLSELYGLDDSTNKHVPYTVTFBR 893
Db 841 GFTPRFTRWKSGKDVLPATPENDNLYWFRNALKQLLSELYGLDDSTNKHVPYTVTFBR 900
QY 894 SQVRLQHTDSRYVPLWSSVVSRYNHYERTASPPQCSQNTLSSDRGQPLKQLSVQYP 953
Db 901 PQVRLQDGTATASVPLWASVVSRYNHYERTASPPQCSQNTLSSDRGQPLKQLSVQYP 960
QY 954 RQOQPAINLYPDTLPDKLLANSYDDQOQLRALTYYQSSWHHLTNNTVTVGLIPDSTRDI 1013
Db 961 RRQOFTNFPYPTLPTLTFASSYDDQOQLRALTYYQSSWHHLTNNTVTVGLIPDSTRDI 1020
QY 1014 FTYGAENVPAAGLNLLELSDKNLIADDKPREYLGQOKTAYTDGQNTTPTLOTPTTQALIA 1073
Db 1021 FTYDAKQVVDGLNLEALCAENSLIADDKPREYLNQORTFTYTDGKNQAPLEIPTQALIA 1080
QY 1074 FTETTVFNQSTLSAPNGSI PSDKLSSTTLEQAGYQQTNYLPFRPTGDKKVVAAHGTDTYGT 1133
Db 1081 FTETAVLTESLSLAFDGGITPDELPGILTQAGYQOQEPYLPFRPTGDKKVVAAHGTDTYGT 1140
QY 1134 AAOFWRPQOKSNTQLTGKITLIWIDANYCVVQVTRDAAGLTTSKAYDMRFLTPVQLTDLND 1193

Db 1141 EAQFWRPVAQRNTLLTGKTKTLQWDTHYCVITQTQDAAGLTVLVANDWRFLTPVQLTDIND 1200
 Qy 1194 NOHLITLDALGRPTTLRFWGTENGWTCYSSPEKASPPSDVNAATLKKPLPVAOCV 1253
 Db 1201 NVHLITLDALGRPVTRQFENGWVKTYGSSPEKPPSPDINTALTALGPLPVAOCV 1260
 Qy 1254 YAPESWPEVLQKTFNRLAEQDWOKLVNARIITDGRICLTAYRRWQSQKAIPLQISLL 1313
 Db 1261 YAPSWPELFSQEFNTLTQEQTLRLDLRIITDWRICALARRRWLSQASFLPNLL 1320
 Qy 1314 NNGPRLPHSLTLTTDRYDHPDQIQOQVFSDFGRLLQAAARHAGMARQNEGSL 1373
 Db 1321 TNSIGLPHNLTLTPDRYDRDSGQIQHOQVAFSDGFRLLQASVRHEAGAWQRNQGSL 1380
 Qy 1374 IINVOHTENRVAVTGRTEYDNKGPIRTYQYFELNDWRYUSNDSAROKEAYADTHYDP 1433
 Db 1381 VTKMEDTKRWAVTGRTEYDNKGPIRTYQYFELNDWRYUSNDSAROKEAYADTHYDP 1438
 Qy 1434 IGREIKVITAKGWFRRLTFTFWFTVNEDENTAAEV 1469
 Db 1439 IGREIRVITAKGMLRQSQYFFWFTVSEDENTAADV 1474

RESULT 5

AA33730
 ID AA33730 standard; protein; 1481 AA.

AC AA33730;

XX 09-NOV-1999 (first entry)

XX Photorhabdus luminescens 1481 amino acid insecticidal toxin.

XX Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.

XX Photorhabdus luminescens.

XX WO9942589-A2.

XX 26-AUG-1999.

XX 18-FEB-1999; 99WO-EP001015.

XX 20-FEB-1998; 98US-00027080.

XX 20-JAN-1999; 99US-0116439P.

XX (NOVS) NOVARTIS AG.

XX (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.

XX Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MW;
 Chen JS;

XX WPI; 1999-527479/44.

XX N-PSDB; AA206831.

XX New nucleic acid from Photorhabdus luminescens encoding insecticidal
 toxins, used for making resistant transgenic plants.

XX Claim 26; Page 140-145; 148pp; English.

XX This sequence represents a 1481 amino acid insecticidal toxin from
 CC Photorhabdus luminescens. It is one of three insecticidal toxins
 CC (AA33728-Y33730) encoded by open reading frames (orfs) in a 38kb
 CC fragment of P. luminescens DNA (AA206831). This sequence is encoded by
 CC orf2. P. luminescens is a member of the Enterobacteriaceae family and is
 CC a symbiotic bacterium of nematodes of the genus Heterorhabditis. The
 CC nematodes colonise insect larvae, kill them, and their offspring feed on
 CC the dead larvae. However, the insecticidal agents are produced by P.
 CC luminescens rather than the nematodes. The toxins have activity against
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni), European
 CC Corn Borer (Ostrinia nubilalis) and Fall Armyworm (Spodoptera frugiperda)
 CC and also against Coleopteran insects (e.g., Colorado Potato Beetle,
 CC Leptinotarsa decimlineata). In addition the toxins are active against

CC strains resistant to known insecticides. The DNA sequence can be used to
 CC generate transgenic plants of various species that are resistant to
 CC economically important insect pests and also for recombinant production
 CC of the toxins for use as insecticides

XX Sequence 1481 AA;

Query Match 76.5%; Score 6042; DB 2; Length 1481;
 Best Local Similarity 76.1%; Pred. No. 0;
 Matches 112; Conservative 113; Mismatches 226; Indels 14; Gaps 5;

Qy 1 MONSQDSITELSLPKGGGALTGMGEALTTPGPDMAALSPLPISAGRGVAPFTLNYN 60
 Db 1 MONSQTSMTLSLPGGGALTGMGEALTTPAGPDMAALSPLPISAGRGVAPSLTLNYN 60
 Qy 61 SGAGNSPFLGWDQNVMTIRRTHFVPHYDEDTTFLGPGEVLVVA-----DQPRDES 114
 Db 61 SGTGNSPFLGWDQNVMTIRRTSTGVPNYDEDTTFLGPGEVLVVAALNEAGADIRSES 120
 Qy 115 TLQGINLGATFTVGYRSRLESHPFSLRYEQPKTKTDFWLIYSPDQVHLLGKSPQAR 174
 Db 121 SLOGINLGMTFTVGYRSRLESHPFSLRYEQPKTKTDFWLIYSPDQVHLLGKSPQAR 180
 Qy 175 ISNPSOTTQTAQMLLEASVSRSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 234
 Db 181 ISNPLNVNTAQMLLEASVSRSRGEIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVY 240
 Qy 235 GNTASSETLPGLDGSAPSQADWLFYVDFYGERSNLKPFPASTTGSWLCRQRFRIYE 294
 Db 241 GNLTASEVFPFTLNGDDPLKSGWLFCLVFDYGERKNSLSEMPFPKATSNWLCRDRFRIYE 300
 Qy 295 YGFEIRTRRLCRQVLMYHHIQAALDSKITTEHNGPTLVSRLLIINYDESAIASTLVPRVVG 354
 Db 301 YGFAIRTRRLCRQVLMYHHIQAALDSKITTEHNGPTLVSRLLIINYDESAIASTLVPRVVG 360
 Qy 355 EQDG-NVVTLPPLPPELAYQDFSPRHHAHQWPMDFLANFNAIQRWOLVDLKGELPGLLYQD 413
 Db 361 EQDGTVALPPELAYQDFPEPEQKALWRPMDVLANFNTIQRWQLLDLQGEVGVILYQD 420
 Qy 414 KGAWYRSARLGEIGSDAVTWKMQPLSVIPLSQSNASLVLDINGDQLDQVITGPGLRG 473
 Db 421 KNGWYRSARQRTGEENNAVTVGKQQLLPITPAIQDNASLMDINGDQLDQVITGPGLRG 480
 Qy 474 YHSORPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGA 533
 Db 481 YHSQHPDGSWTRFTPLHALPIEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGT 540
 Qy 534 KGKDVVQSGDITLFPVGADPRKLVAFSDVLGSGQAHVLEVSATKVTCPNLRGRFGQPI 593
 Db 541 EGRDVVQSGGTLPLPGADARKLVAFSDVLGSGQAHVLEVSATKVTCPNLRGHGRFGQPI 600
 Qy 594 TLPGFSQPAFEPNPAQVYLADLDSGSGPTDLIYVHTNRLDIPLNKSNGFAEPVTLRPEG 653
 Db 601 TLPGFSQSAANFNPDRLVHLADLDSGSPADLIYVHADHLDFSNESGNGFAQPFTLRFPDG 660
 Qy 654 LRPDHTCOLQWADVQGLGVASLLISVPHMSPHWRCDLTNMPKWLLEMMNNMVGVHTLR 713
 Db 661 LRFDHTCOLQWADVQGLGVVSLISVPHMAPHWRCDLTNAKPNLLEMMNNMGAHTLH 720
 Qy 714 YRSSSQFWLDEKAAALTGTQTPVCYLPFPHTLWQTEDEISGNKLVTLRLYARGAWDG 773
 Db 721 YRSSVQFWLDEKAAALATGTPVCYLPFPVHTLWQTEDEISGNKLVTLRLYARGAWDG 780
 Qy 774 RERFRFGVGVQEDTSHQLACGNAPERTPALTKQWTATGLPVIDNALSTEYWR-DOAF 832
 Db 781 RERFRFGVGVQEDTSHQLACGNAPERTSPALTKQWTATGIPVDNTLSAGYMGDTQAF 840
 Qy 833 AGSPRFTTQDNKQVPLTPEDDMSRYFNRAKGLLRSELVGLDDSTKNKVPYTVTEF 892
 Db 841 TGFTPHFTLWKRGKQVPLTPEDDHNLYWLNALKGQFLRSELVGLDGAQOKIPYTVTES 900
 Qy 893 RSQVRLQHTDSRYFVLWSSVSRNHYERIASDDQCSQNTILSSRDFGQPLKQLSVQY 952

Db 661 RFDDTCQLQVADVOGLGVVSLISVPHMA PHHWRCDLTNAKPWLLSETNNMNGANHTLIH 720
Qy 715 RSSQFWLDEKAAALTTGQTPVCYLPPPIHTLMOTETEDISGNKLVTTLIYARGANDGR 774
Db 721 RSSQFWLDEKAAALTTGQTPVCYLPPPIHTLMOTETEDISGNKLVTTLIYARGANDGR 780
Qy 775 EREFRGYVEQTDHSHOLAQGNAPERTPPALTKWYATGLPVIDNALSTEWYR-DDQAF 833
Db 781 EREFRGYVEQTDHSHOLAQGNAPERTPPALTKWYATGLPVIDNALSTEWYR-DDQAF 840
Qy 834 GFSPRFTWQNDKVPITPEDNRSRYFNRAKQGLLSRLYGLDDSTNKHVPVTVTEFR 893
Db 841 GFTRPFLWKEGKDVPLTPEDDNLNWLNRALKQPLSELYGLDGSAAQOIPYTVIESR 900
Qy 894 SQVRLQHTDSRYVPLWSSVVSRYNHYERIASPQCSQNTLSRDPGQPLKQLSVQYP 953
Db 901 PQVRLQDQATVSPVWASVVSRYNHYERIASPQCSQNTLSRDPGQPLKQLSVQYP 960
Qy 954 RROQPAINLYPDTLPDKLLANSYDDQORQLRLTYQSSWHLLTNTVRLVGLPDSRSDI 1013
Db 961 RNNKPTNYPDITLPTLFASSYDDQORQLRLTYQSSWHLLTNTVRLVGLPDSRSDI 1020
Qy 1014 FTYGAENVPAGLNLLELSDRNSLIADKPREYLGQKQATYTDGQNTTLPLOTPTROALIA 1073
Db 1021 FTYDAKQVPDGLNLETICAENSLIADKPREYLNQRTFTYDCKNQTPKLTPTROALIA 1080
Qy 1074 FTETVFNQSTLSPFNGSIPSDKSLSTLEQAGYQQTNYLPPTRGEDKVVVAHGYTDYGT 1133
Db 1081 FTETAVLTESLSPFNGSIPSDKSLSTLEQAGYQQTNYLPPTRGEDKVVVAHGYTDYGT 1140
Qy 1134 AAQFWRPQKQNTOLTGKTLIWDANVCVVQTRDAAGLTTSYAKYDWRFLTPVOLTIND 1193
Db 1141 EAQFWRPQKQNTOLTGKTLIWDANVCVVQTRDAAGLTTSYAKYDWRFLTPVOLTIND 1200
Qy 1194 NQHLITLDALGRPTLRFWGTENGKMTGYSPKASFPSPDNNAAIELKKPLVAAQCV 1253
Db 1201 NVHLITLDALGRPTLRFWGTENGKMTGYSPKASFPSPDNNAAIELKKPLVAAQCV 1260
Qy 1254 YAPESWMPVLSQKTFNRLAEQDWKLYNARIITEDGRICITLAYRRWQSKAIQPLISLL 1313
Db 1261 YAPESWMPVLSQKTFNRLAEQDWKLYNARIITEDGRICITLAYRRWQSKAIQPLISLL 1320
Qy 1314 NNGPRLPHSLTLTDRYDHPDQOIRQOVVFSDFGRLQLQAARHAGMARQNEGSL 1373
Db 1321 TNSGLPPLHSLTLTDRYDHPDQOIRQOVVFSDFGRLQLQAARHAGMARQNEGSL 1380
Qy 1374 IINYQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHYDP 1433
Db 1381 VTKVENTKTRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAYADTHYDP 1438
Qy 1434 IGRKIVITAKGWFRTLTFTWFTVNEDENTAAE 1468
Db 1439 IGRKIVITAKGWFRTLTFTWFTVNEDENTAAE 1473

RESULT 7

ADR21576
ID ADR21576 standard; protein; 1476 AA.

XX AC ADR21576;
XX DT 04-NOV-2004 (first entry)
XX DE Photorhabdus TcdB1 toxin.
XX KW toxin; insect; insecticidal; transgenic; pest control.
XX OS Photorhabdus luminescens.
XX FN W02004067727-A2.
XX PD 12-AUG-2004.
XX Qy 415 GAWWYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDWVITGFLRGY 474

PF 07-JAN-2004; 2004WO-US0000394.
XX 21-JAN-2003; 2003US-0441723P.
XX PA (DOWC) DOW AGROSCIENCES LLC.
XX PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX NI W, Zhu B, Merlo DJ, Apel-Birkhold PC;
XX WPI; 2004-580999/56.
XX DR N-PSDB; ADR21502.
XX PT Controlling or inhibiting an insect, useful for pest control, comprises
XX PT contacting the insect with effective amounts of a Protein A, a Protein B,
XX PS and a Protein C.
XX Claim 1; SEQ ID NO 22; 368pp; English.

The invention relates to a novel method for controlling or inhibiting an insect comprising contacting the insect with effective amounts of a Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C is encoded by a naturally occurring gene or has an amino acid sequence that differs from the product encoded by a naturally occurring gene only by truncation or by conservative amino acid changes. Protein A is a 230-290 kDa toxin complex insect toxin that is derived from a first taxonomic species, has stand alone insecticidal activity, and has an amino acid sequence at least 40% identical to a sequence selected from XptA1wi, XptA2wi, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin complex potentiator having an amino acid sequence at least 40% identical to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wi, XptB1b, XptB1(orf5), or SepA. Protein C is a 90-120 kDa toxin complex potentiator having an amino acid sequence at least 35% identical to a sequence selected from TcdC1, TcdC2, TcdC3, TcdC4, TcdC5, XptB1wi, XptC1xb, PptC1(orf 6 long), PptC1(orf 6 short), and SepC. Also claimed is a transgenic plant or plant cell that produces a Protein A, a Protein B, and a Protein C. The method is useful for pest control. The present sequence represents Photorhabdus luminescens TcdB1 toxin.

Qy 1 MONSDSITELSLPKGGAITGCGEALTTPGPDGMAALSPLPISAGRGYAPFTLNYN 60
Db 1 MONSDSITELSLPKGGAITGCGEALTTPGPDGMAALSPLPISAGRGYAPFTLNYN 60
Qy 61 SGAGNSPFLGWDGCVNMTIRRRTHFGVPHYDETDITFLGPEGEVLVVA-----DQPRDES 114
Db 61 SGTGNSPFLGWDGCVNMTIRRRTHFGVPHYDETDITFLGPEGEVLVVA-----DQPRDES 120
Qy 115 TLOGINLGATFTVTGYSRSLSHESFSLRYEQPKTTGKTDFTWLIYSPDGQVHLLKSPQAR 174
Db 121 SLQGINLGATFTVTGYSRSLSHESFSLRYEQPKTTGKTDFTWLIYSPDGQVHLLKSPQAR 180
Qy 175 ISNPSTQTTQATWLLASVSRSRGQIYVQYRAEDDTGCEADEITHLQATQRYLHIVY 234
Db 181 ISNPSTQTTQATWLLASVSRSRGQIYVQYRAEDDTGCEADEITHLQATQRYLHIVY 240
Qy 235 GNRTASSETPLGLDGSAPSOADWLFYVDFYGERSNLKTTPAFSTTGSWLCRQDRFSRYE 294
Db 241 GNLTASDVFTPLNGDDPLKSGWMECLVFDYGERKNSLSEMPFKATGNWLCRQDRFSRYE 300
Qy 295 YGFEIRTRRLCRQVLMYHHLQALDSKITENHPTLVSRLLINLYDESAITSLVFRVRVGH 354
Db 301 YGFEIRTRRLCRQVLMYHHLQALDSKITENHPTLVSRLLINLYDESAITSLVFRVRVGH 360
Qy 355 EQDGNVVTLPLELAYQDFSPRRHAWQPMVDVLNFAIQRWQLVDLKGELPGLLYODK 414
Db 361 EDNNTVTALPPELAYQDFSPRRHAWQPMVDVLNFAIQRWQLVDLKGELPGLLYODK 420
Qy 415 GAWWYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDWVITGFLRGY 474

```

421 NGWYRSQRAQGEEMNAVTVGKQWLLPITPAVDNANSLMDINGDQGLDWITGPGLRGY 480
475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIQPKSVRLYANTRDGFAP 534
481 HSQHPDGSWTRFTPLHALPIEYSHPRAQLADLMGAGLSDLVLIQPKSVRLYVNRDGFTE 540
535 GKDVQSGDITLTPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTWPNLGRGFRGQBIT 594
541 GRDVQSGDITLPLPGADARKLVAFSDVLGSGQAHLEVSATQVTCWPNLGHGRFGQBIT 600
595 LPGSQPATEPNPAQVYLADLGGPTDLIVVHTWELDIPLNKGSGFAEPVTLFFPGL 654
601 LPGFSQSAASPNPVRVHLADLGGSPADLIYVHADRLDIFNKGSGFAKPTTISFPDGL 660
655 RFDHTCOLQADVQGLGASVILSPHMSPHHRCDLTNMKPMLNENNNNGVHHTLRY 714
661 RPDTCQLQADVQGLGVVSLISVPHMAPHHRCDLTNAPWLLSETNNNGANHTLHY 720
715 RSSQFVLDKAAALATGQTPVCYLPPPIHTLMOTETEDISGNKLVTLRIYARGAMDGR 774
721 RSSQFVLDKAAALATGQTPVCYLPPVHTLMOTETEDISGNKLVTLRIYARGAMDGR 780
775 EREFRGFYVQTSQSHQLAQGNAPERTPPALTKWYATGLPVIDNALSTEWYR-DDQAPA 833
781 EREFRGFYVQTSQSHQLAQGNAPERTPPALTKWYATGLPVIDNALSTEWYR-DDQAPA 840
834 GFSPREFTWQNDKVPILPEDNRSYFNWRALKGQLLSELYGLDSTNKHVPYVTVTEFR 893
841 GFTPRFTLWKGKGVLPILPEDDNDHNLWLNRLKGOPLSELYGLDGSAAQQQIPYTVTESR 900
894 SQVRLQHTDSRYPYLWSSVVSRYHYERIASDPQCSQNTLSSDRFGQPLKQSVQYP 953
901 PQVQLQDQATVSPVLMASVVSRYHYERIASDPQCSQNTLSSDRFGQPLKQSVQYP 960
954 RROQPAINLYPDTLPDKLLANSYDDQQLRLTTLVQSSWHHLTNVTVLGLPSTRSDI 1013
961 RNRKPTTPYPTLPTDPLTFASSYDDQQLRLTTLVQSSWHHLTNVTVLGLPSTRSDA 1020
1014 FTYGAEVNPAGLNLLELSDKNSLIADKPREYLGQOKTAVTDGQNTPLPTPQALIA 1073
1021 FTYDAKVPVQGLNLETCAENSLIADKPREYLNQRTFTYDGNQPLKTPQALIA 1080
1074 FTTFTVFNQSLTSAFNGSIPSKLSTLLEAGYQQTNYLFPRTGEDKVVVAHGYTDYGT 1133
1081 FTTAVLTSLTSAFNGSIPSKLSTLLEAGYQQTNYLFPRTGEDKVVVAHGYTDYGT 1140
1134 AAQFWRPQKQNTQLTGKILTIWANYCVVQTRDAAGLTTSAYKDWRFLLTPVQLTDIND 1193
1141 EAQFWRPVQARNSLTGTWTLKWDTHYCVITQTDAAGLTVSANYDWRFLLTPVQLTDIND 1200
1194 NOHLITLDALGRPTLTFWGTENGWMTGYSSPEKASRSPSDVNAAIEKKPLPVAQCOV 1253
1201 NVHLITLDALGRPTVQRFWGTENGWMTGYSSPEKASRSPSDVNAAIEKKPLPVAQCLV 1260
1254 YAPESWMPVLSQKTFNRLAEQDWKLVNARIITEDGRCTTLAYRRWQSKAIPQLISLL 1313
1261 YAPDSWMPVLSQKTFNRLAEQDWKLVNARIITEDGRCTTLAYRRWQSKAIPQLISLL 1320
1314 NNGRPLPHSLTTLTDRVDHPPEQIRQOVVPSDGFGRLLQAAARHAGMARQNEGSL 1373
1321 TNSIGLPHSLTTLTDRVDHPPEQIRQOVVPSDGFGRLLQAAARHAGMARQNEGSL 1380
1374 IINVOHTENRVAWVGTERTYDNKQPIRTYQVYELNDWRYVYVNDGARQKEAYATHVYDP 1433
1381 VTKRNTKTRVAWVGTERTYDNKQPIRTYQVYELNDWRYVYVNDGARQKEAYATHVYDP 1438
1434 IGREIKVITAGWFRFTLTFWFTVNEDENTAAE 1468
1439 IGREIRVITAGWFRFTLTFWFTVNEDENTAAD 1473

```

```

ABM70526
XX ABM70526 standard; protein; 1486 AA.
AC ABM70526;
XX
XX
XX
XX 20-NOV-2003 (first entry)
XX
XX Photorhabdus luminescens protein sequence #3623.
XX
XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
XX detection; food; gene expression; plant; animal; microorganism; toxin;
XX antibiotic; biopesticide; virulence factor; disease model; plague;
XX whooping cough.
XX
XX Photorhabdus luminescens.
XX
XX WO200294867-A2.
XX
XX 28-NOV-2002.
XX
XX 07-FEB-2002; 2002WO-IB003040.
XX
XX 07-FEB-2001; 2001FR-00001659.
XX
XX (INSP ) INST PASTEUR.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX
XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX Buchrieser C;
XX
XX WPI; 2003-148459/14.
XX
XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
XX Claim 2; SEQ ID NO 3623; 1205pp; French.
XX
XX The invention relates to the isolation of genes and their encoded
XX proteins from Photorhabdus luminescens. The isolated sequences are
XX sources of probes and primers for detecting the genome of P. luminescens
XX and related species; to study polymorphisms; for gene analysis and for
XX detection/amplification of the genes. Antibodies (Ab) raised against the
XX polypeptides encoded by the genes are used for detection/identification
XX of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
XX carry a gene-containing vector are used to select compounds that
XX modulate, regulate, induce or inhibit expression of the genes in plants,
XX animals or microorganisms other than P. luminescens and are able to alter
XX response or sensitivity to toxins and antibiotics produced by P.
XX luminescens. Cells transformed to express the genes are useful for
XX recombinant production of the proteins, particularly toxins and
XX antibacterials useful as insecticides, bactericides and fungicides. The
XX genes, proteins, vectors containing the genes and Ab are also useful
XX therapeutically (to treat microbial infection by bacteria or fungi that
XX are sensitive to P. luminescens-encoded toxins or antibiotics) and as
XX biopesticides. Other uses of the genes and the proteins are as virulence
XX factors and for identifying targets of human diseases for which P.
XX luminescens is a model (particularly plague and whooping cough). This
XX sequence represents one of the isolated P. luminescens proteins
XX
XX Sequence 1486 AA;
XX
XX Query Match 58.6%; Score 4627.5; DB 6; Length 1486;
XX Best Local Similarity 58.4%; Pred. No. 0;
XX Matches 869; Conservative 223; Mismatches 368; Indels 29; Gaps 11;

```

```

Qy 1 MNSQDSFISLTLSPKGGGATGGEALTPTGPGMAALSPLPISAGRGAPATLYNN 60
Dd 1 MQDSFEVSTLTLSPKGGGATGGEALTPTGPGMAALSPLPISAGRGAPATLYNN 60
Qy 61 SGAGNSPFLGQDCNVMTIRRTTHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDE 113
Dd 61 SSAGNGFFGIGWQCGVMTISRTHGIPQYGNDDTFLSPQGEVNNIALNQGPDIRQDV 120

```


Db 899 TVSESRYQVRISIPVNKETELSAWVTAIENRSYHYERIITDPQPSQIKLQHDIFGSLQS 958

XX

DR	WPI; 1998-179427/16.	QY	651	PEGLRFDHTCOLQADVOGLGASLILSVPHMSPHHWRCDLTNNKPMWLNENNNMGVHH	710
DR	N-FSDB; AAV2924.	Db	660	PEGVQFNTCOLQVADIQGLGASLILTVPHIAHHWRCDLSLTKPMLNWNNGGAH	719
PT	Isolated toxins from Photorhabdus luminescens strains - useful for	QY	711	TLRYSSSQFWLDEKAAALTTGTPVCLYPPPIHTLMQTEDEISGNKLVTTLYARGA	770
XX	control of insect pests.	Db	720	TLHYSSAQFWLDEKQLTRAGKSPACVLPFPMHLLWYETIQDEISGNRLTSEVNSHG	779
PS	Claim 34; Page 196-200; 321pp; English.	QY	771	WDGERFERGVEYBQTDHSHOLAQGNAPERTPPALTKNMYATGLPVDINALSTEVNR	829
CC	The present sequence represents a protein named TcaC of the bacterium	Db	780	WDGERFERGVEYBQTDHSHOLAQGNAPERTPPALTKNMYATGLPVDINALSTEVNR	839
CC	Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the	QY	830	QAFAGSPFRFTWQ--DNKDVLPTEPDNNSRYWFRNALKGQLLRSELYGLDSTNKHVPY	887
CC	nematodes of the Heterorhabdus genus. The bacterium has at least 4	Db	840	QAFAGSPFRFTWQ--DNKDVLPTEPDNNSRYWFRNALKGQLLRSELYGLDSTNKHVPY	898
CC	distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are	QY	888	TVTFRFRQVRRLQHTDSRYVPLVSSVVSRYHYERIASDPQCQNITLSSDRFQPLKQ	947
CC	produced from these regions that are associated with insecticidal	Db	899	TVTFRFRQVRRLQHTDSRYVPLVSSVVSRYHYERIASDPQCQNITLSSDRFQPLKQ	958
CC	activity. The native toxins are secreted proteins. The proteins are toxic	QY	948	LSVQYPRQPAHLYPDTLPDKLLANSYDDQQLRLTYQSSWHHTLNNTVRVGLPD	1007
CC	to insects upon exposure and especially when ingested. The nucleic acid	Db	959	LSVQYPRQPAHLYPDTLPDKLLANSYDDQQLRLTYQSSWHHTLNNTVRVGLPD	1018
CC	sequence can be used to produce transgenic plants, baculoviruses or	QY	1008	STRDIFTYGAENVPAAGLNLLELLSDKNLSIADDPREYLQOQKTAAYTDGQNTTLPQTPT	1067
CC	microbial hosts for toxin production. They can be used to control insects	Db	1019	STRDIFTYGAENVPAAGLNLLELLSDKNLSIADDPREYLQOQKTAAYTDGQNTTLPQTPT	1078
CC	pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera,	QY	1068	RQALIAFTTFTVFNQSTLAFNGSPISDKLSTTLEAGYQQTNYLFPRTGDKVVAHNG	1127
CC	Diptera, Acarina or Homoptera orders, especially the Southern or	Db	1079	RQALIAFTTFTVFNQSTLAFNGSPISDKLSTTLEAGYQQTNYLFPRTGDKVVAHNG	1138
CC	Western corn rootworm, Colorado potato beetle, mealworm, boll weevil,	QY	1128	YTDYGTAAQFWRPQKSNQTLTKITLWDANYCVVOTRDAAGITTSKAYDWRFLTPVQ	1187
CC	turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth,	Db	1139	YTDYGTAAQFWRPQKSNQTLTKITLWDANYCVVOTRDAAGITTSKAYDWRFLTPVQ	1198
CC	corn earworm, European corn borer or tobacco hornworm or budworm	QY	1188	LTDINDNQHILTLDALGRPITLRFWGTENGKWTGYSSPEKASFPPSDVNAAILKPLP	1247
XX	Sequence 1485 AA;	Db	1199	LTDINDNQHILTLDALGRPITLRFWGTENGKWTGYSSPEKASFPPSDVNAAILKPLP	1255
QY	Query Match 58.0%; Score 4581.5; DB 2; Length 1485;	QY	1248	VAQCVVAVESWMPVLSQKTFNRL---AEQWQKLYNARIITEDGRICITLAYRWVQSK	1304
QY	Best Local Similarity 57.6%; Pred. No. 0;	Db	1256	VAQCVVAVESWMPVLSQKTFNRL---AEQWQKLYNARIITEDGRICITLAYRWVQSK	1315
QY	Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;	QY	1305	AIPLIISLNNNGPRLPEHSLTLTDRYDHPDQQLRQOVVFSDFGRLQLAAARHAGWA	1364
QY	1 MONSDRSITELSLPKGGGALTGCEALTPTGPDGMAALSPLPISACRGVAPFTLNN	Db	1316	AIPLIISLNNNGPRLPEHSLTLTDRYDHPDQQLRQOVVFSDFGRLQLAAARHAGWA	1375
Db	1 MQDSPEVSIITLSPKGGGALTGCEALTPTGPDGMAALSPLPISACRGVAPFTLNN	QY	1365	RQNEDEGLLIINVQ-----HTENRWAVTGRTEYDNKGQPIRTYQYFLNDWRYVNSDA	1418
QY	61 SCAGNSPFLGWCNVMVTRRTHGPHYDPTDFLQPEGEVLVA-----DQPRDE	Db	1376	RQNEDEGLLIINVQ-----HTENRWAVTGRTEYDNKGQPIRTYQYFLNDWRYVNSDA	1435
Db	61 NSAGNPPGIGWQCGWMSISRTTGHGIPQYGNDDTFLSPQGEVNNIALNDQGPDRQV	QY	1419	ROEKAYADTHVYDPIGRIKVIITAKGWFRRLTFTFWFTVNEDENDTAA	1467
QY	114 STLOGINLCAFTVGYRSLRSHESFSLRYWQPKT--TGKTDWLYSPDGOVHLGSP	Db	1436	ROEKAYADTHVYDPIGRIKVIITAKGWFRRLTFTFWFTVNEDENDTAA	1482
Db	121 KTLQGVTLPISTVTVRYQALIDFSKLEYWQPAQSGQSGRA-FWLIISTPDGHLHLGKTA	QY	AAW17887	AAW17887 standard; protein; 1485 AA.	
QY	172 QARLSNPQTQTQAWLLEASVSSRGEQIYQYRAEDDTGCEADITHLQATQRYLHI	Db	AAW17887	AAW17887 (revised)	
Db	180 QACLANPQNDQIAQWLEETVTPAGEHVSQYRAEDAHCCDNEKTAHPNVTAQRYLVQ	QY	AC	AAW17887;	
QY	232 VYGNRTASSETPLGLDGSAPQADWLFYLVFDYGYRSNNLTPPAFST-TGSWLCRQRP	Db	DT	17-OCT-2003	
Db	240 VYGNRIKQASLFLVDNAPPAPPEWLFHLDHGERDTSHTVPTWADAGTAQWSVRPDI	QY	DT	29-JAN-1998	
QY	291 SRYEYGFIRTRRLCROVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESALSTLVFVR	Db	XX	Photorhabdus luminescens insect toxin protein TcaC.	
Db	300 SRYEYGFIRTRRLCROVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESALSTLVFVR	QY	XX	Insecticide; insect; toxin; pest control; biological control;	
QY	351 RVGHQDGNVTLPLLELAYQDFSPRHAAWQPMQDLANFNAIQRWQVLDLKGEGPLGL	Db	XX	Photorhabdus luminescens; TcaC; Southern corn rootworm;	
Db	360 QLSHESDRPVTQPLELAWQFDEKFTWQFDALONFNSQRYQLVLDLKGEGPLGL	QY	XX	Colorado potato beetle; Western corn rootworm; meal worm; boll weevil;	
QY	411 YQDGAWYRSARLGEIGSDAVTWKMQPISVIPSLSNASLVNDINGDGLDWWITGPG	Db	XX	turf grub; Coleoptera; beet armyworm; black cutworm; cabbage looper;	
Db	420 YQDGAWYRSARLGEIGSDAVTWKMQPISVIPSLSNASLVNDINGDGLDWWITGPG	QY	XX	codling moth; European corn borer; tobacco hornworm;	
QY	471 LRGHVSQRPDGSWTFPLNALPVEYTHPRQALADLMGAGLSDLVLIIGPKSVRLYANTRD	Db	XX		
Db	480 LRGHVSQRPDGSWTFPLNALPVEYTHPRQALADLMGAGLSDLVLIIGPKSVRLYANTRD	QY	XX		
QY	531 GFAKGVVQSGDITPLPGADPRKLVAFSVGLSGQALHVEVSATKVTCPNLGRFRG	Db	XX		
Db	540 GWRKEDVPQSTGTLPTVTGTDARKLVAFSVGLSGQALHVEVSATKVTCPNLGRFRG	QY	XX		
QY	591 QPITLPGSQPATEFNPAQVYLDLGGSPDPLIIVHNRDLDFLNKSGNGPABVPTLRF	Db	XX		
Db	600 QPITLPGSQPATEFNPAQVYLDLGGSPDPLIIVHNRDLDFLNKSGNGPABVPTLRF	QY	XX		

Db 1133 PQISDVAGSEKVVVARQGYTEYGAAQYRFLIQKSLITGKYLTDWTHYCVVVKTED 1192
 Qy 1169 AAGLTTSKADYWRFTLPVQLTDINDNOHLITDAGRPTTLRFWGTENGKMTGYSSPEKA 1228
 Db 1193 GAGMTTQAKYDTRFLPQAULTDINDNOHLITVFNALGOVTSRFRWGTENGKISGYTPESK 1252
 Qy 1229 SPSPPDVNAATELKPLPVAQCVYAPESMMPVLSQKTFN---RLAQDQWQKLYNARI 1285
 Db 1253 PFTVPDTEKALALQPTIPVSCNIVVPDSWMRLLPQQSLTQQLKEGETLNNALHRAGVV 1312
 Qy 1286 FEGDICTLAYERWVQSKAIPQ----LISLLANGPRLPPLSHLTLTDRYDHDPPQOIRO 1341
 Db 1313 FEGDICTLAYERWVQSKAIPQ----LISLLANGPRLPPLSHLTLTDRYDHDPPQOIRO 1372
 Qy 1342 OVVFSDGFRLLQAAARHEAGMARONEDGSLIIN-----YQHTENRWAIVTGRTEYDNK 1395
 Db 1373 SILVSDGFRVQLQSRHEAGMARONEDGSLIIN-----YQHTENRWAIVTGRTEYDNK 1432
 Qy 1396 GQIRTYQYFYFNDWRYVNSDAROEKEAYADTHVYDPIGRBKIVITAKMFRRLTFPM 1455
 Db 1433 GQAIRAYLPYYLNDWRYVNSDAROEKEAYADTHVYDPIGRBKIVITAKMFRRLTFPM 1490
 Qy 1456 FTVNEDENDTAAEV 1469
 Db 1491 FTVNEDENDTAAEV 1504

RESULT 13
 ID AEB47812 standard; protein; 1506 AA.
 AC AEB47812;
 XX

DT 22-SEP-2005 (first entry)
 XX

DE Native XptB1 (xb).
 XX

KW XptB1; toxin; insect resistance; insecticide.
 XX

OS Xenorhabdus bovienii.
 XX

PN US2005155104-A1.
 XX

PD 14-JUL-2005.
 XX

PF 23-DEC-2004; 2004US-00020849.
 XX

PR 07-JAN-2004; 2004US-0534893P.
 XX

PA (APEL/) APEL-BIRKHOOLD P C.
 PA

PA (HEYT/) HEY T D.
 PA

PA (THOM/) THOMPSON R L.
 PA

PA (MEAD/) MEADE T.
 PA

PA (LIZS/) LI Z S.
 PA

PA (RUSS/) RUSSELL S M.
 PA

PA (SHEE/) SHEETS J J.
 PA

PA (LIRA/) LIRA J M.
 PA

PA (FENC/) FENCIL K J.
 PA

PA (MITC/) MITCHELL J C.
 XX

PI ApeI-Birkhold PC, Hey TD, Thompson RL, Meade T, Li ZS;
 PI

PI Russell SM, Sheets JJ, Lira JM, Fencil KJ, Mitchell JC;
 XX

DR WPI; 2005-496874/50.
 XX

DR N-PSDB; AEB47811, AEB47817, AEB47819.
 XX

XX New isolated protein and encoding nucleic acid having toxin activity
 PT against an insect, useful for developing new insecticidal toxins that can
 PT be used to control insects.
 XX
 PS Claim 3; SEQ ID NO 2; 54pp; English.
 XX
 CC This sequence represents native XptB1 (xb). XptB1 is one of five TC (toxin

CC complex) proteins from *Xenorhabdus*. XptA1 is a "stand alone" toxin. XptA2
 CC also has some stand alone toxin activity. XptB1 and XptC1 are the
 CC *Xenorhabdus* potentators that can enhance the activity of either (or
 CC both) of the XptA toxins. XptD1 has some level of homology with TccB
 CC (toxin complex c protein B). This protein has toxin activity against an
 CC insect. The polynucleotide encoding this protein may be used to transform
 CC a plant cell, and thereby generate a transgenic plant which may be
 CC ingested by an insect, whose numbers are subsequently controlled. The
 CC methods and compositions of the present invention are useful for
 CC developing new insecticidal toxins and other proteins that can be used to
 CC control insects.
 XX
 SQ Sequence 1506 AA;

Query Match 53.9%; Score 4261; DB 9; Length 1506;
 Best Local Similarity 54.1%; Pred. No. 0;
 Matches 819; Conservative 228; Mismatches 409; Indels 58; Gaps 21;
 Qy 2 QNSQFSTELSLPKGGGAIQMGAEALPTGPDGMAALSLPLPISAGRGYAFATLYNYS 61
 Db 3 QDSQDMVTYQLSLPKGGGAIQMGDTISNAGPDGMAALSLPLPISAGRGYAFATLYNYS 62
 Qy 62 GAGNSPFLGMDNCVMTIRRTTHGCVPHYDETDTLGPEGEVLVVA---DOP--RDEST 115
 Db 63 GAGNSPFLGMDNCVMTIRRTTHGCVPHYDETDTLGPEGEVLVVA---DOP--RDEST 122
 Qy 116 LOGINLGATFTVGYRSRLESFSLRYWQPKT-TGKTDFMLIYSPDGQVHLGLKSPQAR 174
 Db 123 LLGGQLPVYTVTRHOPRNIQHFSKLEYWQPTDVTETPFMLMYSPDQIHIFGKTEQAQ 182
 Qy 175 ISNPSQTTQATQWLLLEASVSRGEIYYQYRAEDDTGCEADITHHLQATQRYLHIVY 234
 Db 183 IANPAEVSQLAQWLLLEETVTPAGEHIYYQYRAEDDTGCEADITHHLQATQRYLHIVY 242
 Qy 235 GNRFTSETPLGLDGSAPSQADWLFYLVFDYGCERSNNLKTTPAF-STGSGMLCQRQFSRV 293
 Db 243 GNRFTSETPLGLDGSAPSQADWLFYLVFDYGCERSNNLKTTPAF-STGSGMLCQRQFSRV 302
 Qy 294 EYGFIRTRRLCRQVLMYHHLQAL-DSKITEHNGPTLVSRILNYDESAIATLVFVRV 352
 Db 303 EYGFIRTRRLCRQVLMYHHLQAL-DSKITEHNGPTLVSRILNYDESAIATLVFVRV 362
 Qy 353 GHEQGNVVTLPPLLEAYQDFSPRHHAWQPMVDVLANFNAIORWQLVDLKGGLPQLLYQ 412
 Db 363 AYETDAILPPLLEAYQDFSPRHHAWQPMVDVLANFNAIORWQLVDLKGGLPQLLYQ 422
 Qy 413 DK-GAWYRSARQLGEIGS-DAVTEKMQPLSVIPSLQSNASLVINDGQDLDWITGPG 470
 Db 423 DRPGAWYQAPIRQKNVEDINAVTSPINLPKIPSQQDRATLMDIDGGHLDWITAGAG 482
 Qy 471 LRGHYSQRPDGSWTRFTPLNALPVEYTHPRQAQADLMAGLSDLVLIIGPKSVRLYANTRD 530
 Db 483 LRGHYSQRPDGSWTRFTPLNALPVEYTHPRQAQADLMAGLSDLVLIIGPKSVRLYANTRD 542
 Qy 531 GFAGKQVQSGDITLPVPGADPRKLFVPSVILGSGQAHLEVSATKVTWNLGRGRG 590
 Db 543 NWKAGINWMPDGVNLPFGGDSALVAFSDMLGSGQAHLEVSATKVTWNLGRGRG 602
 Qy 591 QPITLPGESQAPATEENPAQVYLADLDGSGPTDLIYVHTNRLDI FLNKSNGFAEPVTLRF 650
 Db 603 AAILPFGSPNGTFTNANQVFLADLDGSGPTDLIYVHTNRLDI FLNKSNGFAEPVTLRF 662
 Qy 651 PEGRLFDHTCOLQADVQGLGVASLILSVPHMSPHHRCOLTNMKPMLNEMNNMGVHH 710
 Db 663 PEGVNFDTCCQLQVSDIOGLGAASIVLTPHMTPRHRYDFTTHKFWLLNVNNRGAST 722
 Qy 711 TLRYRSSQFWMLEKAAALTTGQTPVCYLPFPFIHTLMOTETEDEISGNKLVTLRYARGA 770
 Db 723 TLRYRSSQFWMLEKAAALTTGQTPVCYLPFPFIHTLMOTETEDEISGNKLVTLRYARGA 782
 Qy 771 WDGRERFRFGYVEQTSQHLAQGNAPERTP-----PALTKNWTATGLPVDNALSTSEY 825
 Db 783 WDGRERFRFGYVEQTSQHLAQGNAPERTP-----PALTKNWTATGLPVDNALSTSEY 841


```

Qy 713 RYRSSQWLDKAAALTTGTPVCYLPPFHTLMQTETEDEISGNKLVTLRYARGAWD 772
Db 720 HYRSSQWLDKAAALAGSSPACLYLPTLTLWRSVVQDEITGNRLVSDVLYRHGWD 779
Qy 773 GREERFGVVEQTDHOLA-QGNAPERTPALTQWYATGLPVIDNALSTEWYRDD-Q 830
Db 780 GQERFGRFGVEIRDTDTLASQGTATLSTLSPSVSRNMYATGVPAVDERLPETWQDAA 839
Qy 831 AFAGFSRFTTWQDNKQVLPEDDNRYSRNFNRALKGOLLSELYGLDDSTNKHVPVTVT 890
Db 840 APADPAFTVGSSEDEQTYTP-DDSKTFWLQRLKGLLSELYGADGSSQADIPSVT 898
Qy 891 EFRSQVRLQHTDSRYPLWMSVSVESRNYHYERIASDPQCSQNTLSSDRFGQPLKQLSV 950
Db 899 ESREPQV-RLVEANGDPYVWPMGAESRTSVYRHYNDPQCQQAALLSDRYGFLRQVSV 957
Qy 951 QYPRQQAINLYPDTLPDKLLANSYDDQQLRLTVQSSWHLLTNTVTR-VLGLPDS 1008
Db 958 NYPRPPSADNPYPASLPATLFANSYDQQLRLGLQSSAHLVLSSEGHLLGLAEA 1017
Qy 1009 TRSDIFTYGAENVAGGLNELLSDKNSLIADDPREVLGQKQYATYDQNTTTPLOTPT 1068
Db 1018 SRDVFYTSADNVPEGGLTLEHLLAPESLVSQVGLAGQQVWYLDSDQVATVAPPL 1077
Qy 1069 QALIAFTETTVFNQSTLSAFNGSPSKLSTLLEQAGYQNTYLP--RTGEDKVVVAHH 1126
Db 1078 PPKVAFIETAVLDEGVSSLAAYIVDEH---LEQAGYRQSGYLPFRGREAEQALWTCQ 1133
Qy 1127 GYTDYGTAAQFWRQKQNTOLTKITLIWDANYVVVQTDAAGLTTSACKYDRELTPV 1186
Db 1134 GYVTYAGAEHFWLPLSRDMLTGPVTVTRDAYDCVITQWQDAGIVTTADYDWRFTPV 1193
Qy 1187 QLTIDINDOHLITLDALCRPITLRFWGTENGKMTGYSPKASPPSDVNAALIELKKPL 1246
Db 1194 RVTDPNQLQSVTLDALGRVTLTFWGTENGIAIGYSD---ATLSVPDGAALALATPL 1250
Qy 1247 PVAQCQVAPESWNVLSQKTFNRLAEQDWOKLYNARIITEDGRICITLAYERVVOQKAI 1306
Db 1251 PVAQCLVVTDSW-----GDDNKE----- 1270
Qy 1307 POLLSLNNRPLPHSLTLTDYDHPDPEQOIRQOIVFSPGFGELLQAAARHEAGMARQ 1366
Db 1271 -----MPHVVVVLTADRYSDTGQOVRQOQVTFSDGFGRELQSATRQAGNAWQ 1318
Qy 1367 RNEDGSLI-----INVQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSAR 1419
Db 1319 RGRGKLVASDGLPVTVV-TNFRWAVTGRAEYDNKGLPVRYQPYFLDSQWYVSDSAR 1377
Qy 1420 QEKAYADHYVDYPIGRKIVITAKGWPRRTLFTFPWFTVNEDENDT 1465
Db 1378 Q--DLIADTHFYDPTAREWQVITAKGERQVLYTFWFWVSEDENDT 1421

RESULT 15
ADR21540
ID ADR21540 standard; protein; 1428 AA.
XX AC ADR21540;
XX AC ADR21540;
XX DT 04-NOV-2004 (first entry)
XX DE Serratia SepB toxin SEQ ID NO:60.
XX KW toxin; insect; insecticidal; transgenic; pest control.
XX OS Serratia entomophila.
XX PN WO2004067727-A2.
XX PD 12-AUG-2004.
XX PF 07-JAN-2004; 2004WO-US000394.

```

```

XX 21-JAN-2003; 2003US-0441723P.
PR (DOWC) DOW AGROSCIENCES LLC.
PA Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
XX WPI; 2004-580999/56.
DR Controlling or inhibiting an insect, useful for pest control, comprises
PT contacting the insect with effective amounts of a Protein A, a Protein B,
PT and a Protein C.
XX Claim 1; SEQ ID NO 60; 369pp; English.
XX The invention relates to a novel method for controlling or inhibiting an
CC insect comprising contacting the insect with effective amounts of a
CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
CC is encoded by a naturally occurring gene or has an amino acid sequence
CC that differs from the product encoded by a naturally occurring gene only
CC by truncation or by conservative amino acid changes. Protein A is a 230-
CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic
CC species, has stand alone insecticidal activity, and has an amino acid
CC sequence at least 40% identical to a sequence selected from XptAlwi,
CC XpCA2w1, TcdA, TcdA2, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin
CC complex potentiator having an amino acid sequence at least 40% identical
CC to a sequence selected from TcdB1, TcdB2, TcdC, TcdC1w1, XptB1xb,
CC PptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
CC having an amino acid sequence at least 35% identical to a sequence
CC selected from TcdC1, TcdC2, TcdC3, TcdC4, TcdC5, XptB1w1, XptC1xb, PptC1
CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic
CC plant or plant cell that produces a Protein A, a Protein B, and a Protein
CC C. The method is useful for pest control. The present sequence represents
CC Serratia entomophila SepB toxin.
XX Sequence 1428 AA;

```

```

Query Match 50.3%; Score 3971; DB 8; Length 1428;
Best Local Similarity 52.7%; Pred. No. 2.3e-309;
Matches 783; Conservative 186; Mismatches 431; Indels 86; Gaps 16;

Qy 1 MONSQDSITSLSPKGGGALTGGEALTTPGPDGMAALSPLPISAGRGVAPFTLVN 60
Db 1 MONHQDAITAPTLPSCGGAVTGLKIDAAAGPAAATLSPLPSPGRGVPAGTALNYH 60
Qy 61 SGAGNSPFLGWCNVMTIRRTTHFGVPHYDETDFLCPEGEVLVA-----DQPRDES 114
Db 61 SRSGNGPFGIOWGIGGAQVQRTRNGAPTVDYDTDFGPDGEVLVPALTAAQTQEARQT 120
Qy 115 TLQGINLGATFTVTGYSRLESFHSRLLEYWQPKTTGKTDFWLIYSPDGQVHLLGKSPQAR 174
Db 121 SLGLINPGGSFNVQVYRSRTSGLSRLERWLPADETETEFWVLYTPDQVALLGRNAQR 180
Qy 175 ISNPSQTTQATQWLLLEASVSSRGQIYYQYRAEDDTGCEADEITHLQATQRYLHIVY 234
Db 181 ISNPTAPTQATVAVLWLMSSVSLTGQMYQYQYRAEDDDGCEADEAHPQAGRYPAVWY 240
Qy 235 GNRRTASSETLPGLDGSAPSOADWLVFLVDFYDERSNNLKTTPAFSTTGS--WLCRODFSR 292
Db 241 GNRQAARTLPAL-VSTFSDMSWLFILVDFYDERSVLSSEAPAWQTPGSGEWLQRCDFSG 299
Qy 293 YEYGFETIRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLINYDESAIATSLVFRVR 352
Db 300 YEPGFNLTRRLCRQVLMFHYLVGLAGSSGANDAPALISRLLLDYRESPSLLENVHQV 359
Qy 353 GHEQDGNVVTLPPLLEYAQDPSRHHAHQWQMDVLNPNFNAQRLQVLDLKGELPGLLYQ 412
Db 360 AYESDGTSCALPALALAGWQTFPTPTLSAWQTRDDMGKLSLLQPYQLVDLNGEGVGVILYQ 419
Qy 413 DKGAWYRSQRLGEIGSDAVTWKMQPLSVIPSLQSNASLVNDINGQDLQDWITPGRL 472
Db 420 DSGAWTYREPVRQSGDDPDAVTWGAAALPTWPAHLNHSGLLADLNGDLRLEWVTAQVA 479

```


GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:45:19 ; Search time 133.704 Seconds
(without alignments)
4606.314 Million cell updates/sec

Title: US-10-754-115-45
Perfect score: 7901
Sequence: 1 MNSQDPSITSLSPKGGGA.....WFTVNEEDNTAEVKKVKM 1474

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA_Main:
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	US-10-706-424-10	Sequence 10, Appl
2	7901	100.0	1474	US-10-754-115-45	Sequence 45, Appl
3	6031.5	76.3	1476	US-09-817-514A-4	Sequence 4, Appl
4	4581.5	58.0	1485	US-10-262-794A-32	Sequence 32, Appl
5	4261	53.9	1506	US-10-754-115-49	Sequence 49, Appl
6	4261	53.9	1506	US-11-020-848-2	Sequence 2, Appl
7	3971	50.3	1428	US-10-754-115-60	Sequence 60, Appl
8	3834	48.5	1493	US-10-753-901-18	Sequence 18, Appl
9	3834	48.5	1493	US-10-754-115-18	Sequence 18, Appl
10	3106.5	39.3	1444	US-10-609-113-11	Sequence 11, Appl
11	3106.5	39.3	1444	US-10-754-115-40	Sequence 40, Appl
12	1262	16.0	697	US-10-609-113-39	Sequence 39, Appl
13	202	2.6	2386	US-10-156-761-7751	Sequence 7751, Ap
14	199	2.5	2060	US-10-381-596A-2	Sequence 2, Appl
15	181.5	2.3	2364	US-10-156-761-7834	Sequence 7834, Ap
16	162.5	2.1	1250	US-10-156-761-7572	Sequence 7572, A
17	160.5	2.0	1385	US-10-283-122A-68242	Sequence 68242, A
18	160.5	2.0	1426	US-09-912-020-340	Sequence 340, App
19	160.5	2.0	1426	US-10-282-122A-42617	Sequence 42617, A
20	160.5	2.0	1426	US-10-771-241-340	Sequence 340, App
21	160.5	2.0	1551	US-10-437-963-181412	Sequence 181412,
22	159.5	2.0	1329	US-10-450-763-56139	Sequence 56139, A
23	159.5	2.0	1329	US-10-450-763-58761	Sequence 58761, A
24	154	1.9	5215	US-09-861-289-2	Sequence 2, Appli
25	154	1.9	5215	US-09-860-846-2	Sequence 2, Appli
26	154	1.9	5215	US-09-988-384B-2	Sequence 2, Appli
27	154	1.9	5215	US-09-836-821-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-706-424-10 5215 4 US-10-271-889-45 Sequence 45, Appl
; Sequence 10, Application US/10706424 Sequence 20637, A
; Publication No. US20040103455A1 Sequence 21, Appl
; GENERAL INFORMATION: Sequence 13, Appl
; APPLICANT: ffrrench-Constant, Richard Sequence 14, Appl
; APPLICANT: Waterfield, Nicholas Sequence 157, App
; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens Sequence 1323, Ap
; FILE REFERENCE: 62878 Sequence 12, Appl
; CURRENT APPLICATION NUMBER: US/10/706.424 Sequence 93, Appl
; CURRENT FILING DATE: 2003-11-12 Sequence 5292, Ap
; NUMBER OF SEQ ID NOS: 16 Sequence 53336, A
; SOFTWARE: PatentIn version 3.0 Sequence 46731, A
; SEQ ID NO 10 Sequence 43060, A
; LENGTH: 1474 Sequence 61631, A
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-706-424-10

Query Match 100.0%; Score 7901; DB 4; Length 1474;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MNSQDPSITSLSPKGGGALTGMGEALTPTGPDGMAALSLPLPISAGRGVAPAFLLN 60
Db	1	MNSQDPSITSLSPKGGGALTGMGEALTPTGPDGMAALSLPLPISAGRGVAPAFLLN 60
Qy	61	SGAGNSPFLGWCDCNVMTIRRTTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120
Db	61	SGAGNSPFLGWCDCNVMTIRRTTHFGVPHYDETDTFLGPEGEVLVADQPRDESTLQGIN 120
Qy	121	LGATFTVTGYSRLESHPESRLEYWQPTGKTDFWLIYSPDQVHLLGKSPQARISNPSQ 180
Db	121	LGATFTVTGYSRLESHPESRLEYWQPTGKTDFWLIYSPDQVHLLGKSPQARISNPSQ 180
Qy	181	TTQTAQWLLEASVSRRGEQIYQVRAEDDTGCEADEITHLQATQRYLHVYVGNRTAS 240
Db	181	TTQTAQWLLEASVSRRGEQIYQVRAEDDTGCEADEITHLQATQRYLHVYVGNRTAS 240
Qy	241	ETLPGLDGSAPSQADWLFLYVDFYDYGERSNNLKTTPAFSTTGSWLCRQDRFSRYEYGFIR 300
Db	241	ETLPGLDGSAPSQADWLFLYVDFYDYGERSNNLKTTPAFSTTGSWLCRQDRFSRYEYGFIR 300
Qy	301	TRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINDESAIASTLVFVRVGHQDGNV 360
Db	301	TRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINDESAIASTLVFVRVGHQDGNV 360
Qy	361	VTLPPLELAYODFSRRHHAHQPMQMDVLANFNAIQRWQLVDLKGEGLLYQDKGAWWYR 420
Db	361	VTLPPLELAYODFSRRHHAHQPMQMDVLANFNAIQRWQLVDLKGEGLLYQDKGAWWYR 420

```
Qy 421 SAQRLGIGSDAVTWKMOPLSVIPSLQSNASLVNDINGDGLDWMVITGPGLRGVHSORPD 480
Db 421 SAQRLGIGSDAVTWKMOPLSVIPSLQSNASLVNDINGDGLDWMVITGPGLRGVHSORPD 480
Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Qy 541 SGDTLPLVPGADPRKLVAFSDVLGSGQAHLEVSAATKVTWPNIGRGRFGQPIITLPGFSQ 600
Db 541 SGDTLPLVPGADPRKLVAFSDVLGSGQAHLEVSAATKVTWPNIGRGRFGQPIITLPGFSQ 600
Qy 601 PATEFNAQVYVADLDSGPTDLIVHTNRLDIFLNKSGNGFAPVTLRPEGLRFPDHTC 660
Db 601 PATEFNAQVYVADLDSGPTDLIVHTNRLDIFLNKSGNGFAPVTLRPEGLRFPDHTC 660
Qy 661 QLOWADVQGLGVASLILSVPHMSPHWRCDLTNNKPMLLNEMNNMNGVHHTLRYRSSQF 720
Db 661 QLOWADVQGLGVASLILSVPHMSPHWRCDLTNNKPMLLNEMNNMNGVHHTLRYRSSQF 720
Qy 721 WLDEKAAALTGTQTPVCVLPPIHTLWQTEDEISGNKLVTTILRYARGAWDGRERFRG 780
Db 721 WLDEKAAALTGTQTPVCVLPPIHTLWQTEDEISGNKLVTTILRYARGAWDGRERFRG 780
Qy 781 FGYYEQTDSHQLAQGNAPERTPPALTKNWYATGLPVIDNALSTEYWRDQAFAFSPRFT 840
Db 781 FGYYEQTDSHQLAQGNAPERTPPALTKNWYATGLPVIDNALSTEYWRDQAFAFSPRFT 840
Qy 841 TWQNDKDVPLTPEDDNRKYNFNALKGQLLRSELYGLDDSTNKHPVTVTFEFSQVRRLQ 900
Db 841 TWQNDKDVPLTPEDDNRKYNFNALKGQLLRSELYGLDDSTNKHPVTVTFEFSQVRRLQ 900
Qy 901 HTDSRYVPLMSSVVEYRNHYERIASDPQCSQNTILSSDRFGQPLKOLSVQYPRRQPAI 960
Db 901 HTDSRYVPLMSSVVEYRNHYERIASDPQCSQNTILSSDRFGQPLKOLSVQYPRRQPAI 960
Qy 961 NLVYEDTLDPDKLLANSYDDQKQLRLLTYQSSSWHLLTNTVRLVGLPDSSTRSDIFTYGAEN 1020
Db 961 NLVYEDTLDPDKLLANSYDDQKQLRLLTYQSSSWHLLTNTVRLVGLPDSSTRSDIFTYGAEN 1020
Qy 1021 VPAGLNLLELSDKNSLIADDKPREYLGQCKTAYTDGQNTTPTQTPRQALIAFTETTVF 1080
Db 1021 VPAGLNLLELSDKNSLIADDKPREYLGQCKTAYTDGQNTTPTQTPRQALIAFTETTVF 1080
Qy 1081 NQSTLSAFNGSIPSDKLSLTLEQAGYQCTNLYFPRTGEDKVVVAHHGYTDYGTAAQFWRP 1140
Db 1081 NQSTLSAFNGSIPSDKLSLTLEQAGYQCTNLYFPRTGEDKVVVAHHGYTDYGTAAQFWRP 1140
Qy 1141 QKQNTQLTGKILTIWDANYCVVQTRDAAGLTTSYAKYDWRFLTPVQLTDINDNQHLITL 1200
Db 1141 QKQNTQLTGKILTIWDANYCVVQTRDAAGLTTSYAKYDWRFLTPVQLTDINDNQHLITL 1200
Qy 1201 DALGRPITLRFWGTENGKMTGYSPEKASFPSPDNNAAIELKKPLFVAQCQVYAPESWM 1260
Db 1201 DALGRPITLRFWGTENGKMTGYSPEKASFPSPDNNAAIELKKPLFVAQCQVYAPESWM 1260
Qy 1261 PVLQKTFNRLAEQDWKLYNARIITEDGRICTLAYRRVQSQKAIQOLISLNNNGPRLP 1320
Db 1261 PVLQKTFNRLAEQDWKLYNARIITEDGRICTLAYRRVQSQKAIQOLISLNNNGPRLP 1320
Qy 1321 PHSITLTTDRYDHDPEQOIRQOVVFSQGRGLLOAAARHAGMARQORNEGSLIINVQHT 1380
Db 1321 PHSITLTTDRYDHDPEQOIRQOVVFSQGRGLLOAAARHAGMARQORNEGSLIINVQHT 1380
Qy 1381 ENRWAVTGRTEYDNKGQPIRTYQYFNLNDWRYVNSDSARQEKAYADTHVYDPIGREIKV 1440
Db 1381 ENRWAVTGRTEYDNKGQPIRTYQYFNLNDWRYVNSDSARQEKAYADTHVYDPIGREIKV 1440
Qy 1441 ITAKGNFRRTLFTPWFTVNEDENDTAAEVKVKM 1474
Db 1441 ITAKGNFRRTLFTPWFTVNEDENDTAAEVKVKM 1474
```

```
RESULT 2
US-10-754-115-45
; Sequence 45, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 45
; LENGTH: 1474
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-754-115-45

Query Match 100.0%; Score 7901; DB 4; Length 1474;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MONSQDSITELSLPKGGGAIQMGEGALTPGPDGMAALSPLPISAGRGYAPAFITLNYN 60
Db 1 MONSQDSITELSLPKGGGAIQMGEGALTPGPDGMAALSPLPISAGRGYAPAFITLNYN 60
Qy 61 SGAGNSPFLGWDGNCNVTIRRRTHFGVPHYDETDITFLGPEGEVLVWADQPRDESTLOGIN 120
Db 61 SGAGNSPFLGWDGNCNVTIRRRTHFGVPHYDETDITFLGPEGEVLVWADQPRDESTLOGIN 120
Qy 121 LGATFTVTGYSRLESFHSRLEWQPKTKCTKDFWLIYSPDGOVHLLGKSPQARISNPSQ 180
Db 121 LGATFTVTGYSRLESFHSRLEWQPKTKCTKDFWLIYSPDGOVHLLGKSPQARISNPSQ 180
Qy 181 TTQTAQWLLLEASVSSRGEQIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVVYGNRTAS 240
Db 181 TTQTAQWLLLEASVSSRGEQIYYQYRAEDDTGCEADEITHHLQATAQRYLHIVVYGNRTAS 240
Qy 241 ETLPGLDGSAPSQADWLFLVDFDYGERSNNLKTTPPAFTTGSWLCRODRFSRYEYGEFIR 300
Db 241 ETLPGLDGSAPSQADWLFLVDFDYGERSNNLKTTPPAFTTGSWLCRODRFSRYEYGEFIR 300
Qy 301 TRRLCROVLMYHILQALDSKITEHNGPTLVSRLLINLYDESAIATSTLVFVRVVGHEQDQNV 360
Db 301 TRRLCROVLMYHILQALDSKITEHNGPTLVSRLLINLYDESAIATSTLVFVRVVGHEQDQNV 360
Qy 361 VTLPPLLELAYQDPSPRHHAHWQMDVLANFNALQORWOLVDLKGEGFLGLLYQDKGAWMYR 420
Db 361 VTLPPLLELAYQDPSPRHHAHWQMDVLANFNALQORWOLVDLKGEGFLGLLYQDKGAWMYR 420
Qy 421 SAORLIGIGSDAVTWKMOPLSVIPSLQSNASLVNDINGDGLDWMVITGPGLRGVHSORPD 480
Db 421 SAORLIGIGSDAVTWKMOPLSVIPSLQSNASLVNDINGDGLDWMVITGPGLRGVHSORPD 480
Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLIQPKSVRLYANTRDGFAGKDVQV 540
Qy 541 SGDTLPLVPGADPRKLVAFSDVLGSGQAHLEVSAATKVTWPNIGRGRFGQPIITLPGFSQ 600
Db 541 SGDTLPLVPGADPRKLVAFSDVLGSGQAHLEVSAATKVTWPNIGRGRFGQPIITLPGFSQ 600
```

Db 541 SGDIITLVPFGADPRKLVAFSDVLSGGQAHLEVSATKVTWPNLGRGRFGQPIITLPGRSQ 600
Qy 601 PATEFNPAAQVYLADLGGSDPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPEGLFHDTC 660
Db 601 PATEFNPAAQVYLADLGGSDPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPEGLFHDTC 660
Qy 661 QLOMADVQGLGVASLILSVPHMSPHHRCDLTNMKPWLLENMNNMNVHHTLRYRSSQF 720
Db 661 QLOMADVQGLGVASLILSVPHMSPHHRCDLTNMKPWLLENMNNMNVHHTLRYRSSQF 720
Qy 721 WLDEKAAALTTGOTPVCLYPPPIHTLWOTETEDLSGNKLVTTLRYARGAMDGRERPRG 780
Db 721 WLDEKAAALTTGOTPVCLYPPPIHTLWOTETEDLSGNKLVTTLRYARGAMDGRERPRG 780
Qy 781 FGYYQTDSSHQAQGNAPERTPPALTKNWYATGLPVIDNALSTEWYRDDQAFAGSPRFT 840
Db 781 FGYYQTDSSHQAQGNAPERTPPALTKNWYATGLPVIDNALSTEWYRDDQAFAGSPRFT 840
Qy 841 TWQDNKDVPLTPEDDNSRYWENRALKGQLLRSELYGLDDSTNKHVPYVTEFRSQVRRLQ 900
Db 841 TWQDNKDVPLTPEDDNSRYWENRALKGQLLRSELYGLDDSTNKHVPYVTEFRSQVRRLQ 900
Qy 901 HTDSRYPVLMSSVVSERNYHIERIASDPCSONITLSSDRFCQPLKQLSVQVPRQOPAI 960
Db 901 HTDSRYPVLMSSVVSERNYHIERIASDPCSONITLSSDRFCQPLKQLSVQVPRQOPAI 960
Qy 961 NLYPDTLPDKLLANSYDDQORLRTYQOSSWHHLTNNTVRVLGLPDSRSDIFTYGAEN 1020
Db 961 NLYPDTLPDKLLANSYDDQORLRTYQOSSWHHLTNNTVRVLGLPDSRSDIFTYGAEN 1020
Qy 1021 VPAGGLNLELLSDKNSLIADKPRLYLGOQKATAYTDGQNTTLPQTPTQOALIAFTETTVF 1080
Db 1021 VPAGGLNLELLSDKNSLIADKPRLYLGOQKATAYTDGQNTTLPQTPTQOALIAFTETTVF 1080
Qy 1081 NQSTLSAFNGSLPSDKLSTLEAQGYQOTNYLFPRTGDKVVAHGYTDYGTAAQFWRP 1140
Db 1081 NQSTLSAFNGSLPSDKLSTLEAQGYQOTNYLFPRTGDKVVAHGYTDYGTAAQFWRP 1140
Qy 1141 QKQNTQLTKITLIWDANYCVVQTRDAAGLTTSKDYDRFLTPVQLTDINDNQHLLITL 1200
Db 1141 QKQNTQLTKITLIWDANYCVVQTRDAAGLTTSKDYDRFLTPVQLTDINDNQHLLITL 1200
Qy 1201 DALGRPITLRFWGTENGKWTGYSSEKASFPSPSDVNAALIELKPLPVAQCQVAPESWM 1260
Db 1201 DALGRPITLRFWGTENGKWTGYSSEKASFPSPSDVNAALIELKPLPVAQCQVAPESWM 1260
Qy 1261 PVLQKTFNRLAEQDWQKLYNARIITEDGRICTLAYRWVSOKAIPOLISLNNNGRPLP 1320
Db 1261 PVLQKTFNRLAEQDWQKLYNARIITEDGRICTLAYRWVSOKAIPOLISLNNNGRPLP 1320
Qy 1321 PHSLTLTTRDYHDPEQOIROQVVFSDGFRLLQAAARHEAGMARQNRNEDGSLIINVQHT 1380
Db 1321 PHSLTLTTRDYHDPEQOIROQVVFSDGFRLLQAAARHEAGMARQNRNEDGSLIINVQHT 1380
Qy 1381 ENRWAVTORTEYDNKGQPIRTYQVPLNDWRYVNSDSARQEKAYADTHVDPGIREIKV 1440
Db 1381 ENRWAVTORTEYDNKGQPIRTYQVPLNDWRYVNSDSARQEKAYADTHVDPGIREIKV 1440
Qy 1441 ITAKGWFRRTLTFTPWFTVNEDENDTAAAEVKVKVM 1474
Db 1441 ITAKGWFRRTLTFTPWFTVNEDENDTAAAEVKVKVM 1474

RESULT 3
US-09-817-514A-4
; Sequence 4, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS

; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1476
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-4

Query Match 76.3%; Score 6031.5; DB 3; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;

Qy 1 MONQDSFISITELSLPKGGGAIITGMGEALTPTGPDGMAALSPLPISAGRGVAPFTLNYN 60
Db 1 MONQDTFSVTLSLUPKGGGAIITGMGEALTPTGPDGMAALSPLPISAGRGVAPFTLNYN 60
Qy 61 SGAGNSPFLGWDNCNMTIRRTTHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDS 114
Db 61 SGTGNSPFLGWDNCNMTIRRTTHFGVPHYDETDTFLGPEGEVLVVA-----DQPRDS 114
Qy 115 TLQGINLGATPTVTGYRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLR 174
Db 121 SLQGINLGATPTVTGYRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLRSLR 180
Qy 175 ISNPSQTTQTAQMLLEASVSSRGQIYYQYRAEDDTGCEADEITHHLQATQRYLHVY 234
Db 181 ISNPNVNQTAQMLLEASISHSSEQIYYQYRAEDDEAGCETDELAHAHPSATVQYLTQTHY 240
Qy 235 GNRTASETLPGLDGSAPSQADWLFLYFVGYERSNNLKTTPAFSTTGSWLCRQDRFSRYE 294
Db 241 GNLTSADVFTPLNGDDPLKSGMFCFLVFDYGERKNSLSEMLPFKATGNLWCKRDFSRYE 300
Qy 295 YGFEIRTRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAIASTLVFRRVGH 354
Db 301 YGFEIRTRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIINYDESAIASTLVFRRVGH 360
Qy 355 EQDGNVVTLPLELAYQDPSPRHHAHQPMVLANFNALQWQLVLDLAGEPLGLLYODK 414
Db 361 EDNNTVTALPPLLELAYQDPEPEQATLWQSDMVLNFTIQRWQLLDLAGEVPGVGLYQDR 420
Qy 415 GAWYRSQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDWTITGFLRGY 474
Db 421 NGWYRSQRQAGGEMNAVWGMQLLPITAVQDNASLMDINGDQLDWTITGFLRGY 480
Qy 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAK 534
Db 481 HSQHPDGSWTRFTPLHALPIEYSHPRQAQLADLMGAGLSDLVLIGPKSVRLYANNRDGTE 540
Qy 535 GKDVVQSGDITLVPFGADPRKLVAFSDVLSGGQAHLEVSATKVTWPNLGRGRFGQPIIT 594
Db 541 GRDVVQSGDITLVPFGADPRKLVAFSDVLSGGQAHLEVSATKVTWPNLGRGRFGQPIIT 600
Qy 595 LPGFSQAPATEFNPAAQVYLADLGGSDPTDLIYVHTNRDLIFLNKSGNGFAEPVTLRFPEGL 654
Db 601 LPGFSQSAASFPNDRVHLADLGGSPADLIYVHADRLDI FSNESGNGFAKPTLSFPDGL 660
Qy 655 RFDHTCQLQADVQGLGVASLILSVPHMSPHHRCDLTNMKPWLLENMNNMNVHHTLRY 714
Db 661 RFDHTCQLQADVQGLGVASLILSVPHMSPHHRCDLTNMKPWLLENMNNMNVHHTLRY 720
Qy 715 RSSQFQWLDKAAALTTGOTPVCLYPPPIHTLWOTETEDLSGNKLVTTLRYARGAMDGR 774
Db 721 RSSVQFQWLDKAAALTTGOTPVCLYPPPIHTLWOTETEDLSGNKLVTTLRYARGAMDGR 780
Qy 775 EREPRFGYVEQTDSSHQAQGNAPERTPPALTKNWYATGLPVIDNALSTEWYRDDQAF 833
Db 781 EREPRFGYVEQTDSSHQAQGNAPERTPPALTKNWYATGLPVIDNALSTEWYRDDQAF 840

```

Qy 834 GFSRFTTWOQNDVPLTPEDDNRNRYFNRAKLGOLLRSLEYGLDSTNKHVPVTVTEFR 893
Db 841 GFTFRFTLWKEGKDVPLTPEDDNLNRYFNRAKLGOLLRSLEYGLDSTNKHVPVTVTEFR 900
Qy 894 SQVRLQHTDSRYFVPLWSSVVSERNYHYERIASDPQCSQNTITLSSDRFGQPLKQLSVOYP 953
Db 901 PVRQLQDQATVSPVLMASVVSERNSYHYERIASDPQCSQNTITLSSDRFGQPLKQLSVOYP 960
Qy 954 RRQOPAINLYPDTLPDKLLANSYDDQORQLBLTVQSSWHHLTNTVTVLGLPSTRSDI 1013
Db 961 RNRPTTNPYDPTLPDTLTFASSYDDQQLRLTLTQSSWHHLTNTVTVLGLPSTRSDI 1020
Qy 1014 FTYGAENVAPGGLNELLSLDRKSLIADDPREYLGQOQATVTDGQNTPLTPTQALIA 1073
Db 1021 FTYDAKVVDGLNLETCAENSLIADDPREYLGQOQATVTDGQNTPLTPTQALIA 1080
Qy 1074 FTETTFVFNQSLTSAFNGSPISDKLSTLLEQAGYQOQTNLYFPRTGDKVVAHGHYTDYGT 1133
Db 1081 FTETAVLTESLSAFDGGITPDELPGILTQAGYQOQTNLYFPRTGDKVVAHGHYTDYGT 1140
Qy 1134 AAQFWRPQOKQNTOLTGKHTILWANYCVVQTRDAAGLTTSYAKYDRFLTPVQLTIND 1193
Db 1141 EAQFWRPVPAQRNSLTGKHTILWANYCVVQTRDAAGLTTSYAKYDRFLTPVQLTIND 1200
Qy 1194 NOHLITDLCGRPTTLRFWGTENGWMTGYSSPEKASPPSPDVAATLKKPLPVAOCOV 1253
Db 1201 NVHLITDLCGRPTTLRFWGTENGWMTGYSSPEKASPPSPDVAATLKKPLPVAOCOV 1260
Qy 1254 YAPESWMPVLQKTFNRLAEQWQKLYNARIITDGRICITLTYRRVQSQKAIPLQLISLL 1313
Db 1261 YAPDSWMPVLSQETFNRLAEQWQKLYNARIITDGRICITLTYRRVQSQKAIPLQLISLL 1320
Qy 1314 NNGRPLPHSLTLTDRYDHPDQOIRQOQVFSDFGRLLOAAARHEAGMARQNEGSL 1373
Db 1321 TNSIGLPHNLTLTDRYDHPDQOIRQOQVFSDFGRLLOAAARHEAGMARQNEGSL 1380
Qy 1374 IINVOHTENWAVTGRTEYDNGGPIRTYQVYELNDWRYVNSDSARQKAYADTHVYDP 1433
Db 1381 VTKVENTKRWAVTGRTEYDNGGPIRTYQVYELNDWRYVNSDSARQKAYADTHVYDP 1438
Qy 1434 IGREIKVITAKGFRRTLTFTFWFTVNEDENTAAE 1468
Db 1439 IGREIRVITAKGLRQSQYFPFWFTVSEDENTAAD 1473

```

```

RESULT 4
US-10-262-794A-32
; Sequence 32, Application US/10262794A
; Publication No. US20030207806A1
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dow Agrosciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN

```

```

; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/262,794A
; FILING DATE: 02-OCT-2002
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,567
; FILING DATE: 05-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-10-262-794A-32

```

```

Query Match 58.0%; Score 4581.5; DB 4; Length 1485;
Best Local Similarity 57.6%; Pred. No. 0;
Matches 858; Conservative 224; Mismatches 378; Indels 29; Gaps 11;
Qy 1 MONSDPFSITELSLPKGGGALTGMGEALTPTGPGMAALSPLPLISAGRGYAPAFITLNN 60
Db 1 MQDSPEVSITTLSPKGGGALTGMGEALTPTGPGMAALSPLPLISAGRGYAPAFITLNN 60
Qy 61 SGAGNSPFLGWDGNCNMTIRRTTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDE 113
Db 61 NSAGNGPFGIGWQCGVMSISRRTHQGIPOYGNDTFLSPQGEVNMIALNDQGPDIRQDV 120
Qy 114 STLOGINLGATPTTGYRSRLESHPFSLRYWPKT--TGKTDFWLIYSPDGOVHLGKSP 171
Db 121 KTLOGVTLPISTYTRYQARQILDPSKIEYQWQASGOEGRA-FWLITPDGHLHLGKTA 179
Qy 172 QARISNPSQTTQTAQWLLEASVSRGEQIYYQYRAEDDTGCEADEITHHQAQRYLHI 231
Db 180 QACLANPQNDQIAQWLLEETVTPAGEHVSQYRAEDEAHCCDNEKTAHPNVTAQRYLVQ 239
Qy 232 VYVGNRTASETLPLGLDGSAPSQADWLFLVLPDYGYERSNNLKTTPAFST-TGSWLCQRDP 290
Db 240 VNYGNIRPQASFLVLDNAPPAPPEEWLFLVFDHGERDTSLHTVPTWDAGTAQWSVRPDI 299
Qy 291 SRYEYGEIRTRRLCROVLVHHLQALDLSKTEHNGPTLVSRLLIINYDESAIASTLVFVR 350
Db 300 SRYEYGEVTRRLCROVLVHHLQALDLSKTEHNGPTLVSRLLIINYDESAIASTLVFVR 359

```



```
Qy 651 PEGRLFDHTCOLQWADVOGLGASLILSVPHMSPHHWRCDLTNNKFWLLNEMNNMGVHH 710
Db 663 PEGWFDNTCOLQVSDIOGLGASLILSVPHWYDFHNPWLLNINNRGAET 722
Qy 711 TLRVSSQFWLDEKAAALTTGQTPVCYLPPIHPTLMQTEDEISGNKLVTLRYARGA 770
Db 723 TLFYSSAQFWLDEKASQIELGKFAASVLPPIHLLWNEALDEITGNRLTKVWNYAHA 782
Qy 771 WDEREPFGYVEOTDSHQAQNAEPRT-----PALTKWYATGLPVIDNALSTEX 825
Db 783 WDEREPFGYVEOTDSHQAQNAEPRT-----PALTKWYATGLPVIDNALSTEX 841
Qy 826 WR-DQQAFAAGSPRTTW-----QDNKDVPLTPEDDNRYSYFNRALKGQILRSLEYLD 878
Db 842 WRGDDQAFAGTPRTTRYEKNAGQEGQDTIPEPTETEAYWLNAPAMKQQLRSEVYG-D 900
Qy 879 DSTNK-HVPYTVTEPRSVRRLOHTDSRYFVLWSSVVSRYHVERIASDPQCSQNTYLS 937
Db 901 DKTEKAKIPYTVTEARCOVRLIPSNDAAAPSSWTSIIENRSYHYRIIVVDPSCQVVLK 960
Qy 938 SDRGQPLKQLSVQYPRQOQAINLYPDTLPKLLANSYDDOOROLRLTYQSSWHHLTN 997
Db 961 ADEYGFPLAKVDIAIPRNKPAQNPYDPSLPDTLFDADSYDQCKOLYTKQOQSYHHTQ 1020
Qy 998 NTVRLGLPDSRSDIFTY-----GAENVPAAGLNLLELLSDKNSLIADDPREYLGQOKTA 1053
Db 1021 QDDWVLGLTDRYSEVHYAQTDAQSDIPKAGLILELLKVDGLGKDKTYILGQORVA 1080
Qy 1054 YTDGQNTPLQTPTRQALIAFTTTFVFNQSTLSAFNGSIPSDKLTTLLEAGYQOQNTYLF 1113
Db 1081 YVGGD-----AEKPTQRQVAVTETAAPDDNALHAFDGVIAPELTOQLLAGY----LLV 1132
Qy 1114 PRTE-----DKWVAHGYDYGTAAQFWRPQKQNTQLTGKITLIWANDYCVVWQTRD 1168
Db 1133 PQISDVAGSSEKVVWYARQGYTEYGAARQFYRPLIQKSLTGKTYLSWDTHYCVVWKTED 1192
Qy 1169 AAGLTSAKYDWRFLTPVQLTIDINDNHLITLDALGRPITLRFMTGKMTGKTYSSPEKA 1228
Db 1193 GAGMTQAKYDRLPQLDINDNQHIYFNALQVTSRFRWGTENGKISGYSTPEK 1252
Qy 1229 SFSPSDVNAALIELKPLVQAQCVYAPESMMPVLSQKTFN-----RLAEDQWOKLYNARI 1285
Db 1253 PFTVPDTEKALALQPTIPVSCNIYVPSWMLLPQQLSLTGKKEGTLNWLHRAQV 1312
Qy 1286 TEDGRICTLARVWVQSKAIPQ-----LISLLANGPRLPHSLTLTDRYDHDPEQOIRQ 1341
Db 1313 TEDGLICELAYRRWIKRQATSSMMAVTLQQLAQTFRQPPHAMTITTDYDSDSQOQLRQ 1372
Qy 1342 QVWPSDGRLLQAAAREHAGMARONEDGSLIIN-----VOHTENRWAVTGRTEYDNK 1395
Db 1373 SIVLSDFGRVLOSQRHEAGWQRAEDGSLVVDNTGKPVVANTTTWASGRTEYDCK 1432
Qy 1396 QPIRTYQYFINDRWRYNSDSARQEKAYADTHYDPIGRKIVITAKGFRFTLFTPW 1455
Db 1433 GOAIRAYLPYLNDRWRYVSDSARD--DLIADTHFYDPLGREYQVKTAKGFWRENMFPM 1490
Qy 1456 FTVNEDENDTAEEV 1469
Db 1491 FTVNEDENDTAARL 1504
```

RESULT 7

```
US-10-754-115-60
; Sequence 60, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
```

```
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XCI
; CURRENT APPLICATION NUMBER: US/10754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 60
; LENGTH: 1428
; TYPE: PRT
; ORGANISM: Serratia entomophila
; US-10-754-115-60
```

```
Query Match 50.3%, Score 3971; DB 4; Length 1428;
Best Local Similarity 52.7%, Pred. No. 2.4e-304;
Matches 783; Conservative 186; Mismatches 431; Indels 86; Gaps 16;

Qy 1 MONSDPSITSLSPKGGGALTGGEALTPTGPDGMAALSPLPISAGRGYAPATILYN 60
Db 1 MONEHQDWAITAPTLFSGGGAVTGLKGDIAAAGPDGAATLSPLVSPGRGYAPTALNYH 60
Qy 61 SGAGNSPGLGWCDCNVMTIRRTTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDES 114
Db 61 SRSGNGPFGIGWGGGAAVQRRTRNGAPTDTDTFTGPDGEVLVPAUTAAGTQBARQT 120
Qy 115 TLOGINIGATFTVGYRSRLESFHSRLEYWQPKTTGKTDFWLLYSPDQVHLLGKSPQAR 174
Db 121 SLGLINPGGSFNVQYRSRTSGLSRLERWLPADETETEFWVLYTPDGOVALLGRNAQA 180
Qy 175 ISNPSQTTQTAQWLEASVSRGBOIYYQYRAEDDTGCEADBITHHLQATQRYLHYVY 234
Db 181 ISNPTAPTQTAWLMESSVLTGEMYYQYRAEDDDGCEADDAHPQAQRYPVAVY 240
Qy 235 GNRPTASETLGLDGSAPSQAOWLFVLFVFDYDERSNLTTPPAFTTGS--WLCRODRFSR 292
Db 241 GNRQAARTLPAL--VSTPSMOSWLFILVFDYDERSVLSSEAPAWQTPGSGEWLCRODQCSG 299
Qy 293 YEYGFETRRLCRQVLMYHHLQALDSKITSHNGPTLVSRILNYDESAIASTLVFVRV 352
Db 300 YEFGNLTRRLCRQVLMFYHVLGVLGSSGANDAPALISRLLLDYRESPLSLLENVHQV 359
Qy 353 GHEQGNVVTLPPLLEYQDFSPRHHAWQPMQVLANFNAIQRWOLVDLKGELPGLLYQ 412
Db 360 AYESDGTSCALPALAGWQTFPTPTLSAWQTRDDMGKLSLLQPYQLVDLNGEGVGYLYQ 419
Qy 413 DKGAWYRSARLGEIGSDAVTWKMQPLSVIPLSQNASLVLDINGDQOLDWVITGPGLR 472
Db 420 DSGAWYKPEVRPGQDDPDPAVTWGAALPTPALNHSIILADLNGDGRLEWVTPAGVA 479
Qy 473 GYHSORPDGSTRFTPLNALPEVYTHPRAQLADLMAGLSDLVLIGPKSVRLYANTRDGF 532
Db 480 GMYDTEPCRDWLHFTPLSALPEVAHPKAVLADILGAGLTDWVLIGPRSVRLYSKNDGW 539
Qy 533 AKGDVVSQSDITILPVPGADPRKLVAFSDVLGSGQAHLVESVATKVTCPNLRGRFGQP 592
Db 540 NKGETVQQTERTLTPVPGVDPRTLVAFSDMAGSQOHLTEVRANGVRYWPNLGHGRFGQP 599
Qy 593 ITLPGFSQPAFERPAQVYVLADLQSGPTDLIYVHTNRLDILFNKSGNGPPEPVTLRPE 652
Db 600 VNIPIGFSQSVTTFNPDQILLADTDGSGTDLIYAMSRLVIYFNOSGNYFAEPHTLLLPK 659
Qy 653 GLRFDHTCOLQWADVOGLGASLILSVPHMSPHHWRCDLTNNKFWLLNEMNNMGVHHTL 712
Db 660 GVRDRTCSLQVADIQGLGVPSLLLTVPVHAPHWVCHLSADKPWLLNGMNNMGARHAL 719
Qy 713 YRSSSQFWLDEKAAALTTGQTPVCYLPPIHPTLMQTEDEISGNKLVTLRYARGAWD 772
```

```
Db 720 HYRSSVQWFLDKAEALAAAGSSPACLYLFTLTLLWRSVVQDEITGNRLNSDVLRYHGVM 779
Qy 773 GREPEFGFGVEQTDHSHLA-QGNAPERTPPALTKNWYATGLPVIDNALSTEWYRDD-Q 830
Db 780 GQERFPGFGVEIRDTDTLASQGTATSELSMPSVSRNRYATGVPADVPERUPETTYWQDAA 839
Qy 831 AFAGFSPRFTTWQDNKVPLTPEDNSRYWPNRALKGQLRSELYGLDDSTNKHVPYTVT 890
Db 840 AFADPATRFTVGSCEDEQTYTP-DDSKTFWQRALKGILLSELYGADGSSQADIPISVT 898
Qy 891 EFRSQVRLQHTDSRYPVLMSSVVSERNYHYERIASDPQCSQNTLLSDRFGQPLQSLV 950
Db 899 ESRPQV-RLVEANGDYPVWPMGAESRTSVYRVHNDPQCCQQAALLSDEYGFPLRQVSV 957
Qy 951 QYPRQOPAINLYPDTLPDKLLANSYDDQQLBLTVOOSSWHLLTNTVR--VLGLPDS 1008
Db 958 NYPRRPPSADNPYPASLPATLIFANSYDEQQILRLGLQSSAHLVLSLSEGHWLLGLAEA 1017
Qy 1009 TRSDIFTYGAENVPPAGGLNLELLSDKNSLIADDPREVLGQOKTAYTDGQWTTPLQTPTR 1068
Db 1018 SRDVFYTSADNVEGGTLTLEHLAPESLSVDSQVGTLAGQQQWYVLDSDQVATVAAPPL 1077
Qy 1069 QALLAFTETTVFNQSTLSAFNGSIPSDKLSLTLEAGYQQTNYLFP--RTGEDKVVVAHH 1126
Db 1078 PPKVAFIETAVLDSGWSSLAAYIVDEH----LEQAGYRQSGYLFPRGREAEQALWTQCQ 1133
Qy 1127 GYTDYGTAAQWRPQKOSNTOLTKITLIWDANCVVVVQTRDAAGLTSAKYDWRFLTPV 1186
Db 1134 GYVYAGAHEFWLPSFRDSMLTGPVTVTVDAYDCVITQWQDAAGIVTTADYDWRFLTPV 1193
Qy 1187 QLTIDINDNOHLITLIDALGRPTLTFWGTENGKMTGYSSPEKASPPSDVNAALTELKPL 1246
Db 1194 RVDNDNLQSVLIDALGRVTLTFWGTENGIAIYSD---ATLSDVDPGAAALALATPL 1250
Qy 1247 PVAQCQVYAPESMPLVLSQKTFNRLAEQDMQKLYNARIITEDGRICTLAYRRVQSQKAI 1306
Db 1251 PVAQCLVVTDSW-----GDDNDEK----- 1270
Qy 1307 PQLISLNLNGRPLPHSLTLTTRDYDHPDEQIQVQVFSDFGFGRLLOAAARHAGMARQ 1366
Db 1271 -----MPPHVVLATDRYSDTGQVQVQVTFSDGFGRELQSAQTRQAGNAWQ 1318
Qy 1367 RNEDGSLI-----INVQHTENRAWVTRTEYDNKGPDIRTYOPYFLNDWRVYSNDSAR 1419
Db 1319 RGRDGLVTASDGLPVTV-TNFRWATGRABYDNKGLPVRYQPYFLDSHQYVSDDSAR 1377
Qy 1420 QEKEAYADTHYVDPIGREIKVITAKGMFRFTLFTPWFTVNEBENDT 1465
Db 1378 Q--DLXADTHFYDPTAREWQVITAKGERRQVLYTTFWVVSSEBENDT 1421
```

RESULT 8

```
US-10-753-901-18
; Sequence 18, Application US/10753901
; Publication No. US20040194164A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Larrinua, Ignacio
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Schafer, Barry
; APPLICANT: Bevan, Scott
; APPLICANT: Young, Scott
; APPLICANT: Guo, Lining
; TITLE OF INVENTION: Xenorhabdus TC Proteins and Genes for Pest Control
; FILE REFERENCE: DAS-105X
; CURRENT APPLICATION NUMBER: US/10/753,901
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,717
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
```

```
; LENGTH: 1493
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-753-901-18

Query Match 48.5%; Score 3834; DB 4; Length 1493;
Best Local Similarity 50.9%; Pred. No. 1.8e-293;
Matches 767; Conservative 213; Mismatches 472; Indels 56; Gaps 21;

Qy 1 MQNSQDSITELSLPKCGGGAITGMGEALTPTGPDGMAALSILPLISAGRGVAPAFITLNYN 60
Db 1 MQGSTPLKLEIPSLSPCGGSLKGMGEALNAVAGAGGASFLPLPISVGRGLVPVLSLNY 60
Qy 61 SGAGNSPGLGWCDCNVMTIRRRTHFGVPHYDEDTFLGPEGEVL-VVAD---QP--RDES 114
Db 61 STANGSGFGWGCVGFIILRTAKGVPHPTGQDEYLPDGEVLSIVPDSQCGEQRTAT 120
Qy 115 TLOGINIGATFTVTGYSRLESFHSRLYEQP--KTTGKTDFWLIYSPDGOVHLLGKSPQ 172
Db 121 SLLGTVLTPHTVTRYQSRVAEKIVRLEHWQPOQRRREETSFWLFTADGLVHLFGKHH 180
Qy 173 ARISNPQTTQTAQWLLLEASVSRSGEQIYYQYRAEDDTGCEADITHLQATAQRYLIHV 232
Db 181 ARIADPQDETRIARWLMEETVTHTEHIYYHYRAEDDLDCDEHQAQHSQVTAQRYLAKV 240
Qy 233 YYGNRTASETLPLGLDGSAPSOADWLFLVDFDYGERSNNLKTTPAFSTTGS----- 282
Db 241 SYGNTQETAPFAVKSGIPADNDWLFHLVDFDYGERSLNSVPFNVSENNVPEK 300
Qy 283 WLCHQDRFSRYEYGFETRRLCRQVLMYHHLQALDSKITEHNGPTLIVSRILINYDSAI 342
Db 301 WRCHPDSEFSRYEYGFETRRLCRQVLMFHLQALAGEKVAEETPALVSRLLDYDLNNK 360
Qy 343 ASTLVFVRVCHGEDGNVVTLPLELAYQDFSPRHHAHQBMQVLANFNAIQRLQVLDLK 402
Db 361 VSLLTQARLAHAETDGTTPVMSPLEMDYQRVNHGVNLNWQSMQPLEKNTLQPYQLVDLY 420
Qy 403 GEGPLGLLYQD-KGAWMYRSAQR-LGEIGSDAVTWKMQPLSVIPSLQSNASLVDINDGDG 460
Db 421 GEGISGLVLYQDTQAWMYRAPRVDITAEGTNAVYEEAKPLPHIPAQOESAMLLDINDGDG 480
Qy 461 QLDWVITGPGLRGYHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLADLADLADLADLADL 520
Db 481 RLDWVITASGLRGVHTWSPEGWTFFPLSAVPMYEPHPQAKLADIDGAGLPDALIGPN 540
Qy 521 SVRLYANTRGFAKGVQVQSGDITLVPFGADPKLVAFSDVLGSGQAHLVEVSATKVT 580
Db 541 SVRVMSNNRAGWDRAQDVIIHLSDMPLFVPGRNERHLVAFSDMTGSGQSHLVEVTADS 600
Qy 581 WPNLGRGRFGOPITLPGFSQAPATEFNPAQVYVADLDGSGPTDLIYVHTNRLDIPLNKS 640
Db 601 WPNLGRGRFGEPMLMTGFGQISGETFNPDRLYWVDIDSGTDTDFIYARNTYLEYANES 660
Qy 641 GFAEPVTLRFPEGLRFDHTCOLQADYVQGLGVASLILSVPHMSPHHWRCDLTNNKPM 700
Db 661 HFAEPQRIDLPGVRFDDTCRLQIADTQGLGTASIILTIPEMKVQHWRLDMTIFKPML 720
Qy 701 EMNNMGVHHTLYRSSSQFWLDEKAAALTGTQTPVCLPPPIHTLWQTEDEISGNKL 760
Db 721 AVNNMGTTETLYYRSSAQFWLDEKQASESGMTVWSYLPFPVHLVMTRETVLDEISGN 780
Qy 761 VTTLRYARGAWDGRERFRGFGVVEQTDHSHLA---QGNAPERTPPALTKNWYATGLPVI 817
Db 781 TSHVYHSHGAWDGLERFRGFRVGTQDIDRSASATQGTAEPPAPSRVTVWYGTGVREV 840
Qy 818 DNALSTYWR-DDQAFAGFSRFTTWQDNK--DVLPTEDDNSRYWPNRALKGQLRSEL 874
Db 841 DILLPTEYWGQDQAFPHFTPRFTRYDEKSGDGMTVP-SEQEYWLHRLAKGQRLSEL 899
Qy 875 YGLDDSTNKHVPTVTFEFRSQVRELQHTDSRYPVLMSSVVSERNYHYERIASDPQCSQ 934
Db 900 YGDDDSILAGTPYSVDESRTQVRLLPVMSVDPVAVLSVAESRQYRYERVATDQCQSKI 959
```

Qy	935	TLSDRFGQBLKQLSVQYPRRQQAQINLYPDTTLPDKLLANSYDDQQRQLRLTYQQSSWTH	994
Db	960	VLKSDALGFPQDNLEIAYSRPQEPFSPYDPTLPTETLFTSSFEQQMFLRLTQRSSYTH	1019
Qy	995	LT-NNTVRVLGLPDPSTSDIFTTYGAENVPRAGGLNLELLS--DKNSLIADDPKREYLGOOK	1051
Db	1020	LNHDDNTWITGLMDTSTRSDARIYQADKVPDGGFSLEWFSATGAGALLPDDAAADYLGHQR	1079
Qy	1052	TAYTDGQNTTFLQTPTRQALIAFTETTVFNQSTLSAFNGSIPSDKLSLTTLLEQAGYQOTNY	1111
Db	1080	VAYTG-----PEEQAPLPPVAYIETAEPFERSLAAPEEVMDSEQLTKQLNDAGWNTAKV	1134
Qy	1112	LPFRGTGBKVVWAHGHGYDTYGTAAQFWRPQKQNSNTQLTGKITLIWDANYCVVVQTRDAAG	1171
Db	1135	PFSEKTDHFHVMVGQKFEYAGADGFRPLVQRETKLTGKTTVTWDSHYCVIATATEDAAG	1194
Qy	1172	LTTSAKYDREFLTVPQLTDLNDNOHLTLDLALGRPTTLRPMWGTENGWMTGYSSPEKAS--	1229
Db	1195	LRMQAHYDFPMVADNTTVDNDNTHTVTFDALGRVTSFRWGTENGKQGYTFAENETVP	1254
Qy	1230	FSPSPDYNAAIELKPKLPVACQVYAPESWMPVLSSQKTFNRLAEQDWQKLYNARIITEG	1289
Db	1255	FIVPTTVDDALAKPGIPVAGLMVYAPLSNW---VQASFSNDGEL-YGELKPAGITEDG	1310
Qy	1290	RICTLAVRNVQSO--KAIPLISLILNNGPRLPPhSLUTLTTDRYDHDPEQOIQVQVVSFD	1347
Db	1311	YLLSLAFRRWQNNPAAAMPKQVNSQN-----PPHVLSVITDRYDADPEQOLRQTTFSD	1365
Qy	1348	GFGRLLQAAARHEAGMARQNEQSLII-----NVQHTENRVAVTRTEYDKNKGOPIRT	1401
Db	1366	GFGRTLQTAVRHSEGAWVRDEYGAIVAEHNHGAPEATAMTDFRVAWSGRTSYDGKQALRK	1425
Qy	1402	YQPYFLNDRWYVNSDARSQEKAYADTHVYDPIGREIKVITAKGWFRTLTFTPFWTVNED	1461
Db	1426	YQPYFLNSQWYVSDDSARQ--DIVADTHYDPLGREYQVITAKGFFERSLFTPFWVFNED	1483
Qy	1462	ENDTAAEV	1469
Db	1484	ENDTAGEM	1491

RESULT 9

```

US-10-754-115-18
; Sequence 18, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 1493
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-754-115-18

```

Query Match 48.5%; Score 3834; DB 4; Length 1493;
Best Local Similarity 50.9%; Pred. No. 1.8e-293;

Matches	767;	Conservative	213;	Mismatches	472;	Indels	56;	Gaps	21;																																															
Qy	1	MQNSQD	F	S	I	T	E	L	S	L	P	K	G	G	A	I	T	C	M	G	B	A	L	T	P	T	G	D	G	M	A	A	L	S	L	P	I	S	I	S	A	G	R	G	Y	A	P	A	T	I	N	Y	N	60		
Db	1	MQGSTP	L	K	L	K	E	I	P	S	L	P	S	G	G	S	L	K	G	E	A	L	N	A	V	G	A	E	G	G	A	S	F	S	L	P	I	S	I	S	V	G	R	G	L	V	P	L	S	I	N	Y	60			
Qy	61	SGAGNS	P	F	G	L	G	W	C	N	W	T	I	R	R	T	H	F	G	V	P	H	D	E	T	D	T	F	L	G	P	E	G	V	L	--	V	V	A	D	--	O	P	--	R	D	E	S	114							
Db	61	STAGNS	F	G	M	G	W	C	G	V	G	F	I	S	L	R	T	A	K	G	V	P	H	T	G	D	E	Y	L	G	P	D	G	E	V	L	S	I	P	D	S	I	P	D	S	I	P	D	S	I	120					
Qy	115	T	L	O	G	I	N	L	G	A	T	T	V	T	G	R	S	L	E	S	H	F	S	L	E	Y	W	O	P	--	K	T	T	G	K	T	D	F	W	L	I	X	S	P	D	O	V	H	L	G	K	S	P	Q	172	
Db	121	S	L	I	G	T	V	L	T	Q	P	H	T	V	T	R	Q	S	R	V	A	E	K	I	V	L	E	H	W	O	P	O	R	R	E	E	T	S	F	W	L	T	A	D	G	L	V	H	L	F	G	K	H	H	180	
Qy	173	A	R	I	S	N	S	Q	T	T	A	Q	M	L	E	A	S	V	S	G	B	Q	I	Y	Y	R	A	E	D	T	G	C	E	A	D	E	I	T	H	L	Q	A	T	A	O	R	Y	L	H	I	V	232				
Db	181	A	R	I	A	D	Q	O	E	T	R	I	A	R	M	E	E	T	V	T	H	G	E	H	I	Y	H	R	A	E	D	D	L	C	D	E	H	E	L	A	Q	H	S	G	V	T	A	O	R	Y	L	A	K	V	240	
Qy	233	Y	G	N	R	T	A	S	E	T	L	P	G	L	D	G	S	A	P	S	O	A	D	W	L	Y	L	V	P	D	I	G	E	R	S	N	N	L	K	T	P	P	A	F	S	T	G	S	-----	282						
Db	241	S	Y	G	N	T	O	P	E	T	A	F	F	A	V	K	S	G	I	P	A	D	N	D	W	L	F	H	L	V	D	I	G	E	R	S	S	L	N	S	V	P	E	F	N	S	E	N	N	V	P	E	K	300		
Qy	283	W	L	C	R	Q	R	S	R	S	Y	E	G	F	E	I	R	T	R	L	C	R	Q	V	M	H	H	L	Q	A	L	S	K	I	T	E	H	N	G	P	T	L	V	S	R	L	I	N	Y	D	E	S	A	I	342	
Db	301	W	R	C	P	D	S	F	S	Y	E	G	F	E	I	R	T	R	L	C	R	Q	V	M	H	Q	L	A	L	A	G	E	K	V	A	E	T	P	A	L	V	S	R	L	I	D	Y	D	L	N	N	K	360			
Qy	343	A	S	T	L	V	F	R	R	V	G	H	E	Q	D	G	N	V	T	L	P	L	E	A	Y	Q	D	F	S	P	R	H	A	H	M	Q	P	M	D	V	L	A	N	F	N	A	I	O	R	Q	L	V	D	L	K	402
Db	361	V	S	L	L	Q	T	A	R	L	A	H	E	T	D	G	T	P	V	M	S	P	L	E	M	D	Y	Q	R	V	N	H	G	V	N	L	N	Q	S	M	P	Q	L	E	K	M	N									

```
Qy 1052 TAYTDGQNTTPLQTPTRQALIAFTTETTVENOSTLSAENGSPSPDKLSTTLLEQAGYQQTNY 1111
|||
Db 1080 VAYTG-----PEEQPAIPPLVAYIETAEFDRSLAAFEENVDEQELTKQLNDAGWNATKV 1134
|||
Qy 1112 LFPRTGEDKVVAAHHGYTDYGTAAQFWRPQKQSNLTQTKITLTDANDANYCVVVQTRDAAG 1171
|||
Db 1135 PFSEKTDHVVWGQKEFTYAGAGFYRPLVQRETGLTKTITVTDWSDHYCVITATEDAAG 1194
|||
Qy 1172 LTTSKADYWRFLTPVQJLTDINDNOHLITLTDALGRPITLRFWGTENGKMTGYSSEKAS-- 1299
|||
Db 1195 LRMOAHYDYREPMWADNTTVDVNDYHTVTFDALGRVTSPFRFMTENGKQGYTPAENETVP 1254
|||
Qy 1230 FSPSPDVNAATELKPLPVAQCYVAPESWNPVLSOKTENRSLASODWQKLYNARIITEDG 1289
|||
Db 1255 FIVPTTDDALALRPGIPVAGLWYAPLSWM---VQASFSNDGEI-YGELKPAGITIEDG 1310
|||
Qy 1290 RICTLAVRRVQSQ--KAIPQLISLANNPRLPPLPHSLTLTDRYDHPDEQIQOQVVFSD 1347
|||
Db 1311 YLLSLAFRRQNNPAAAMPKQVNSQN-----PHVLSVITDRYDADPEQQLRQTFTFSD 1365
|||
Qy 1348 GFGRLQAAARHEAGMARQNEGSLII-----NVOHTENRWAIVTGRTEVDNKGQPIRT 1401
|||
Db 1366 GFGRTLQATVRHESGEAWVRDEYGAIVAEHNHGAPEATMTDPRWAVSGRTEYDGRKQALRK 1425
|||
Qy 1402 YQPYFLANDWRVVSNDARSQKEAYADTHVYDPIGRBIKVIITAKGFRRTLTFTPMFTVNE 1461
|||
Db 1426 YQPYFLNSWQVSDSARQ--DIYADTHYDPLRGLEYQVITAKGGFRRLSLFTPMFVWNE 1483
|||
Qy 1462 ENDTAAEV 1469
|||
Db 1484 ENDTAGEM 1491
|||

RESULT 10
US-10-609-113-11
; Sequence 11, Application US/10609113
; Publication No. US20040110184A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Bevan, Scott
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from
; FILE REFERENCE: Das-101XC2
; CURRENT APPLICATION NUMBER: US/10/609,113
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US 60/392,633
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/441,647
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 11
; LENGTH: 1444
; TYPE: PRT
; ORGANISM: Paenibacillus strain IDAS 1529
US-10-609-113-11

Query Match 39.3%; Score 3106.5; DB 4; Length 1444;
Best Local Similarity 43.0%; Pred. No. 7.1e-236;
Matches 634; Conservative 230; Mismatches 551; Indels 61; Gaps 17;

Qy 3 NSDPSITELSLPKGGGATGMBALTPGPDGMAALSPLPISAGRGYAPFTLYNSG 62
|||
Db 5 SNADIKLLSPSLPKGGGSMKGIENIAAPGSDGMARCNVLPVTSRGYITPDISLSVAG 64
|||
Qy 63 AGNSPFLGWCNWTMTTTRRRTHFGVPHYDETDFTLGEGEVLVADQPRDE-----STL 116
|||
Db 65 HGNGAYGMGTWGMVNSISRSRTSPSYTSBDQFLGPDGEVLVPESNESQGEIITRHTDTA 124
|||
Qy 117 QGINLGATFTVTGYRSRLSHFSRLEVYQPKTTGKT-DFWLIYSPDGQVHLLGKSPQARI 175
|||
```

```
Db 125 QGIPLGFTFTVTRYPRIESAFLHLEYWEAQAGSATASFWLHISADGVLHCLGKTAQARI 184
|||
Qy 176 SNPSQTTQTAQWLLLEASVSRGEOIYYQYRAEDDTGCEADEITHLQATAQRIYLIHYVG 235
|||
Db 185 AAPDSDAKIAEWLVEESVSPFGEHIYYQYKEEDNQVNLLEDNH--QYGANRYLKSIRY 242
|||
Qy 236 NRTASETLPGLDGSPASQADWLFYLVDFDYGERSNNLKTTPAFSTTGSWLCRQDRFSRYEY 295
|||
Db 243 NKVASPSLYVWKGIBPADGQWLVSVILDYGENDTSDVPPLYTPQGEWLVRPDRFSRYDY 302
|||
Qy 296 GFEIRTRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLIILNYESAISASTLVFVRRVGEH 355
|||
Db 303 GFEVTRCRLCRQVLMFHVKELGGE-----PALVWRMQLDEYDENPAASMLSAVRQLAYE 356
|||
Qy 356 QDGNVVTLPLELAYQDFSPRHHAHWQMPDVLANFNALQRWQLVDLKGEGPLGLLYQDKG 415
|||
Db 357 ADGAIRSLPFLFEDYTFEGTETADWQFLLVPFPEWADDEEHYQLVLDLYGREGIPGLLYQNND 416
|||
Qy 416 AWYRSARQLGEIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDWVITGPLRGYH 475
|||
Db 417 HWHYRSPAR-ODT-PDGIAYNSWRPLPHIPWNSRNGMLMDLNGDCYLEWLLAEFGVAGRY 474
|||
Qy 476 SORPDGCSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGPAKG 535
|||
Db 475 SMNPDKSMSGFVPLQALPTEFFHPQAQLANVTGSLTDLVMIGPKSVRFVAGEAEPKRA 534
|||
Qy 536 KDVVQSGDITLPGADPRKLVAFSDVLGSGOAHLEVSATKYTCWNILGREGFGQITL 595
|||
Db 535 CEVWQVQGITLPVERVDKELVAFSDMLGSGQSHLRIRHGDVTCWNLGNLGVFGAPLAL 594
|||
Qy 596 PGFSQPAEFNPAQVILADLDGSGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRFPEGLR 655
|||
Db 595 HGFTASEREFNPERVILVDLGGSGASDIIYASRDALLIYRLNSGNGFADPRVPLPDGVR 654
|||
Qy 656 FDHTCOLQADVQGLVASLILSVPHMSPHHRWCIDLTMKRPWLNLNMMNNGVHHTLRYR 715
|||
Db 655 FDNLCRLLPADIRGLGVASLVLVHPYMAPRSWKLDFFAAKPYLLQTVSNLNGASSSEFWYR 714
|||
Qy 716 SSSQFWLDEKAAALTTGQTPVCYLPFPIHILWQTEDEISGNKLVTTLTYARCAWDGRE 775
|||
Db 715 SSTQYWLDERQAA-----SSAVSALPFPINVSDMHTVDEISGRTRQTKYTYRHGVYDRT 770
|||
Qy 776 REPRGFGYVQTSQSHQAQGNAPERTPALTKWYATGLPVIDNALSTEVWRDDQAFAG 835
|||
Db 771 KEFAGFRIDTWEERESEGLTSVSTPEVLTTRTYHTGQKQDEERAVQQYWGQDPAFQV 830
|||
Qy 836 SP-RFTTWQ--DNKDVLTPTEDDNSRYWFNRAKLGQLLRSELXGLDDSTNKHVYTVTEF 892
|||
Db 831 KPVELTRFDAAAQDLFLDSNNGQOEYWLVSRLQGMPLRTEIFAGD--VGGSPPYQVESF 888
|||
Qy 893 RSQVRRLOHTDSRYPLWSSVRESNTHYERIASDPOCSQNTILSSRFGOPLKQLSVQY 952
|||
Db 889 RYQVRLVQSIDSECVL-PMQLEQLTYNYEQIASDPOCSQIQOQWDFEYGVAAQSVTIQY 947
|||
Qy 953 PRROQPAINLYPDTLPDKLLANSYDDOOROLRTYQQSSWHLNTNTVRLVGLPDSRSD 1012
|||
Db 948 PRAQPEDNPYPTLPTDTSWSSYDSQOMLLRLTRQKAYHLADPGWRLNIHQITELD 1007
|||
Qy 1013 IFITYGAENVPPAGLNLELLSDKNSLIADDKPREYLQOQKTAITYTDGQNTPTLQTTQAL 1072
|||
Db 1008 AFIVSADSVPAEGISAEELLEVDGTLRSSALEQAQYGGQSEIYAGGGE-----PDLRALV 1061
|||
Qy 1073 AFTETTVFNOSTLSAFNGSIPSDKLSLTLBQAGYQQTNYLFPRTGEDKVVAAHHGYTDY 1132
|||
Db 1062 HYTRASVLDCLQAYEGVLSDSQNLSSASSGYSQRARILGSGDEVDIFVAEGGFTRYA 1121
|||
Qy 1133 TAAQFWEPQKQSNLTQTKITLIWDANYCVVVQTRDAAGLTTSKADYWRFLTPVQLTIDN 1192
|||
Db 1122 DEPNFFRILQOOSLLSGEQVLTWDDNFCAVTSIEDALGNQIQIAYDYRFEALQITDTN 1181
|||
Qy 1193 DNOHLITLTDALGRPITLRFWGTENGKMTGYSSEKASFSPPSDVNAATELKPLPVAQCY 1252
|||
```

Db 1182 NNVAQVALDALGRVVYSRTWGTGEGIKTGF-RPE-VFATPETMEQALALASPLFVASCC 1239
Qy 1253 VYAPESWMPVLSOKTFNRL---AEQDWOKLYNARIITEDGRICITLAYRRWVQSOKAIQOL 1309
Db 1240 VYDAHSWGTITLAQSELVDPSEKQWFLIDNRLIMPDGRIISRGRDPWLSHLRLLPPAV 1299
Qy 1310 ISLNNGRPLPHSLTUTLTDYDHDPEQIQOQVVFSDGFRLLQAAARHAGMARQRNE 1369
Db 1300 GELLSEADKPPHTVILAADRYDDPSQIQASIVFSDGFGRTIQAKRE-----1349
Qy 1370 DGSLLINQHTENRWAVTGRTEYDNKGQPIRTYQVYFLNDMYNSDSARQEKEYADTH 1429
Db 1350 -----DTRWAIAERVDYDGTGAVIRSFQFLLDDNNYVGEAV---SSSMYATII 1396
Qy 1430 VYDPIGRIBKIVITAKGNFRRTLTFTWFTVNEDENDT 1465
Db 1397 YDALARQLRWNAKGYERRTAFYFWFTVNEDENDT 1432

RESULT 11
US-10-754-115-40
; Sequence 40, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 40
; LENGTH: 1444
; TYPE: PRT
; ORGANISM: Paenibacillus strain DAS1529
US-10-754-115-40

Query Match 39.3%; Score 3106.5; DB 4; Length 1444;
Best Local Similarity 43.0%; Pred. No. 7,1e-236;
Matches 634; Conservative 230; Mismatches 551; Indels 61; Gaps 17;

Qy 3 NSQPSITSLPCKGGGAIITGGEALPTGPDGMAALSPLPISAGRGYAPAFILNYSNG 62
Db 5 SNADIKLSPSLKGGGSMKEENIAAPGSDGMARCNVLPVTSGRYITPDISLSYASG 64
Qy 63 AGNSPFLGWCNWNVTIRRHFGVPHYDETDFTLGPGEVLVADQPRDE-----STL 116
Db 65 HNGAYGMGTWGMVMSISRRTSRGTSPYTSDEQFLGPDGVLVPEESNEQGEIIRHTDTA 124
Qy 117 QGINLGAFTVTGYSRLESFHSRLEYWQPTTKGT-DFWLIYSPDGOVHLGKSPQARI 175
Db 125 QGIPUGETFTVTRYPFRLESFHLLEYWEAQAQGSATASFWLIHSGADVHLGCKGTAQARI 184
Qy 176 SNPSOTTQATWLLSEASYSRGEQIYYQYRAEDDTGCBADITHHLQATQRYLHIVYG 235
Db 185 AAPDSAKIAEWLVEESVSPGEHIYYQYKEEDNGVNLLEDNH--QYGANRYLKSIRYG 242
Qy 236 NRTASETLPGLDGAPSQADMFLYLVFYDYGERSNNLKTTPAPFSTGSLCHQDRFSRYEY 295
Db 243 NKVASPSLYWVKGEIPADGQWLYSVILDYGENDTSADVPPLYTPQGEWLVRPDRFSRYDY 302

Qy 296 GFEIRTRRLCRQVLMYHHLQALDQSKITTEHNGPTLVSRILINLYDESAIASTLTVFVRVGH 355
Db 303 GFEVTRTCRLCRQVLMFHFVKELGGE-----PALVWRMQLEYDENPAASMLSAVRQLAYE 356
Qy 356 QDGNVVTLPLELAYQDFSPRHHAHQPMQDVLNPFNAIQRWQLVDLKGEGPLGLYQDKG 415
Db 357 ADGAIRSLUPPLEFDFYTPFGIETTADWQFPLPVEWADEEHYQLVDLYGEGPLGLYQND 416
Qy 416 AWYRSARQLGEISDAVTEWEMQPLSVIPSLQSNASLVDINGDQLDWMVITPGPLRGVH 475
Db 417 HWHYRSPAR-GDT-PDGIAYNSWRPLPHI PVNSRRNGMLMDLNGDGYELWLLAEPGVAGRY 474
Qy 476 SQPDGSWTRTFLNALPVEVTHPRAQADLMAGLSDLVLIGPKSVLYYANTRGDPKAG 535
Db 475 SMNPKSWSGFVPLQALPTEFFHPQAQLANTYVSGSLTDLVMIGPKSVFYPAGEEAGPKRA 534
Qy 536 KDVVQSGDITLVPFGADPRKLVAFDVLGSGOAHLEVSATKVTCPNLRGRGRFOPTIL 595
Db 535 CEVWQVCGITLPUVERVKELVAFSDMLGSGQSHLVRIRHGVTCWPNLGNVGFAPLAL 594
Qy 596 PGFSQPATFNPQAVYLADLDGSGPTDLIYVHTNRLDIFLNKSGNGFAEPVTLRFP 655
Db 595 HGFTASEREFNPERVYLDLGGSGASDIYASRDALLIYRNLGNGFADPVRVPLPDGVR 654
Qy 656 PDHTCOLQMAADVQGLVASLILSVPHMSPHHWRCDLTNMKPWLLNEMNNMGMVHTLRYR 715
Db 655 FDNLCRLLPADIRGLGVASLVHVPYMAPRPSWKLDFFAAKPYLLQTVSNLNGASSFWYR 714
Qy 716 SSSQFWLDEKAAALTGTGTPCYLPPFPIHTLMQTEDEDEISGNLVTTLRYARGADGRE 775
Db 715 SSTQYWLDEKQAA---SSAVSALPFPINNVSDMHTVDEISGRTRTQKYVYRHGYDRTE 770
Qy 776 REFRGFGYVEQTDGSHQLAQGNAPERTPALTKNWTATGLPVIDNALSTEYWRDDQAFAG 835
Db 771 KEPAGFGRIDTWEEERDESEGLTSVSTPPVLTTRTWHTQKQDEERAVQOYWGDDPAAQV 830
Qy 836 SP-RFTTWQ--DNKDVPLTPBEDDNSRYWFNALKGQLRLSELYGLDDSTNKHVPYVTF 892
Db 831 KPVELTRFDAAAQDLPLDSNNGQOEWYVLSRQGMPLRTBIFAGD--VGGSPPYQVESF 888
Qy 893 RSQVRLQHTDSRYPLWSSVVESRNHYEIRASDPQSONITLSSDRFGQPLKLSVQY 952
Db 889 RYQVRLVQSIDSECVL--PMQLEQLTYNYEQIASDPQCSQIQOMFDEYGVAAQSVTIQY 947
Qy 953 PRQOPAINLYPDTLPDKLLANSYDDQORLRLYQSSMWHLTNTNVRVLGLPSTRSD 1012
Db 948 PRRAQPEDNPYRPTLPDTSWSSYDSQWMLRLTRQOKAYHLADPEGRWLNIPHOTRLD 1007
Qy 1013 IFITYGAENVPPAGLNLELLSDKNLSIADDKPREYLQOQKTAITDQONTPTLQTPTRQALI 1072
Db 1008 AFYISADSVPAEGISAELEVDGTLRSSALBQAYGGQSEIIVAGGE-----PDLRALV 1061
Qy 1073 APTETTVNQSTLSAFNGSIPSDKLSLTLEQAGYQOQNTNLYLPTFGEDKVVVAHHGYTDY 1132
Db 1062 HYTRASVLDDECLQAYEGVLSDSQLNSLLASSGYQSRARILGSGDEVDIFVAEQGFTRYA 1121
Qy 1133 TAAQFWRPQKSNQTLGKITLWDANYCVVVQTRDAAGLTTSKAYDWRFITPVOLTIDN 1192
Db 1122 DEPNFRILGQSSLLSGEQVLTWDDNFCVATSTEDALGNQIQIAYDTRFVEAIGITDN 1181
Qy 1193 DNOHLITDALGRPITLRFWGTENGKMTGYGSPKASFPSPSDVNAAILKKPLPVAQCC 1252
Db 1182 NNVAQVALDALGRVVYSRTWGTGEGIKTGF-RPE-VFATPETMEQALALASPLFVASCC 1239
Qy 1253 VYAPESWMPVLSOKTFNRL---AEQDWOKLYNARIITEDGRICITLAYRRWVQSOKAIQOL 1309
Db 1240 VYDAHSWGTITLAQSELVDPSEKQWFLIDNRLIMPDGRIISRGRDPWLSHLRLLPPAV 1299
Qy 1310 ISLNNGRPLPHSLTUTLTDYDHDPEQIQOQVVFSDGFRLLQAAARHAGMARQRNE 1369
Db 1300 GELLSEADKPPHTVILAADRYDDPSQIQASIVFSDGFGRTIQAKRE-----1349
Qy 1370 DGSLLINQHTENRWAVTGRTEYDNKGQPIRTYQVYFLNDMYNSDSARQEKEYADTH 1429

Db 1350 -----DTRWAERVDYDGTGAVIRSFQFFVLDWNNYGEBAV--SSSMYATY 1396
Qy 1430 VYDPIGRBEIKVITAKGWFRTLTFTPWTVNEDENDT 1465
Db 1397 YYDALARQLRMVNAKGYERTAFYFPTVNEDENDT 1432

RESULT 12

US-10-609-113-39
; Sequence 39, Application US/10609113
; Publication No. US20040110184A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Bevan, Scott
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Donald J.
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from
; TITLE OF INVENTION: Paenibacillus Species
; FILE REFERENCE: DAS-101XC2
; CURRENT APPLICATION NUMBER: US/10/609,113
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US 60/392,633
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/441,647
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 39
; LENGTH: 697
; TYPE: PRT
; ORGANISM: Paenibacillus apairius strain DB482
US-10-609-113-39

Query Match 16.0%; Score 1262; DB 4; Length 697;
Best Local Similarity 37.0%; Pred. No. 2.5e-90;
Matches 270; Conservative 119; Mismatches 291; Indels 50; Gaps 11;

Qy 736 VCYLPIPIHLWQETEDDEISGNKLVTTLRYAGAWDGRERFRGFGVEGTDSSHQAQG 795
Db 7 VCGVFPFVNVSDTHTVDEISGSARIQKTYTRNGVYDRTDKEFAGFGHIDTWEEERDSEG 66
Qy 796 NAPTERTPALTKNVAATGLPVIDNALSTEWYRDQAFAGFSP-RFTTWQ--DNKDVPLTP 852
Db 67 TLSISTPPLVTRTHYTHGQKDEERAVQQYWGQDPAFQVQKPVRLTRFDAATQDVFLDS 126
Qy 853 EDDNSRYWFRNALXGOLLRSLEY-----GLDDSTNKHVPYTVTFRSQVRLQHTDSRYP 907
Db 127 PNREEYWLRYSLRGMPLRNEIFAGDVVGLP-----PYQVESLRQVRLMQSTDSECV 179
Qy 908 VLMSSVVESRNYHYERTIASDPQCSQNTLSDDRFGQPLKQLSVQYPRQQAIPNLYPDTL 967
Db 180 TL-PMQLEQLTYNYEQIASDPQCSQIQQWFDEYGVAAQSITI QYPRRAQPEDNYPHTL 238
Qy 968 PDKLLANSYDQOQLALTYQOSSWHHLTNNTVRLGLPOSTRSDIFTYGAEVPAAGLN 1027
Db 239 PDTSSWSSYDSQQMLLRKAYLHTRQKAYHLADPEGWRLNIPHOTRLDSFTYSADSPAEGIS 298
Qy 1028 LELLSDKNSLIADKPREYLGQQQXTAYTDGQNTTPLQTPTRQALIAFTETTVFNQSTLSA 1087
Db 299 AELLGGDTLRS PALEQAYGQSIIYAGGE-----PDSRALVHYTRSAILDEACLOA 352
Qy 1088 FNGSIPSDKJLTLEQAGYQNTYLPFRGTGDKVVAHGHYTDYGTAAQFWRPQKQNTQ 1147
Db 353 YEGVLSQSLNSLLASSGYQSRARILGSGDEADIFVAEQGFTRYADEQNFRIILGQSSSL 412
Qy 1148 LTGKITLIWDANYCVVQVTRDAAGLTTSKAYDWEFLTPVQLTNDNCHLITLDALGRPI 1207
Db 413 LTGEQVLTWDDNFCVATYSIEDALGNQIQIAYDYRFAIRAIQITDANNVNVQSDALGRV 472
Qy 1208 TLRFWGTENGWMTGYSPPEKASFPPSDVNAAIKPKPLPVAQCVQYAPESWMPVLSOKT 1267
Db 473 YSRWTGTEEGTGF-RPE-AEFPSPETMEQALALASPLPVASCVCVDAHSMWGTITLIGQ 530

Qy 1268 FNRL---AEQDWOKLYNARIITEDGRICITLAYRRWVQSKAIPOLISLNNGLPPLPHSL 1324
Db 531 LSAVLPDSEKQWFLIANRLIMPEDGRIRARGRAPWMLQRLLPAPAVAKLLSEADKPPHTV 590
Qy 1325 TLTTRYDHDPEQIIRQOVVFSDFGRLQLQAARHEAGMARQORNEGSLINVOHTENRW 1384
Db 591 VLAADRYPDPSQIQASVVFSDGFGRTIQAKR-----ADTRW 629
Qy 1385 AVTGRTEYDNKGQPIRTYQPYFLNDWRYVNSDSARQEKAEVADTHVVDYDPIGREIKVITAK 1444
Db 630 AITERIDYDETGAVIRSFQFYIDDMNYGKEAV--SGSMYATIIYYDALARQURMWNNAK 687
Qy 1445 GWFRRTLFTP 1454
Db 688 GYERTAFYP 697

RESULT 13

US-10-156-761-7751
; Sequence 7751, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7751
; LENGTH: 2386
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7751

Query Match 2.6%; Score 202; DB 4; Length 2386;
Best Local Similarity 18.3%; Pred. No. 1.7e-05;
Matches 344; Conservative 184; Mismatches 647; Indels 706; Gaps 89;

Qy 13 SLPKGGCAITG--MGEALTPTG---PDGMAALSPLPISAGR---GVAPAFITLNYNS-- 61
Db 318 TLAGSSSDAGDYRASTLSTPTGSWEVSTGSGAFTYVPIQLPKPPMGSAFSLSYNSQS 377
Qy 62 -----GAGN---SPFGLGDCNVMTIRRR-----THFGVPH-----YDDETDFLQPEGEVL 104
Db 378 VDGRTSASNQOASWAGMGWDLNVGYIERRYNCSEDLPTIGDMCWDSPNSAKEPFGAVY 437
Qy 105 VVADQPRDESTLQGINAGTFTVTG---YRSRLSHFSRLEYWQPKTKGT-DFWLIYSP 160
Db 438 VI-----NLNGVTSSELIDQNTGSGAYHLKNDPGWRVQRLFDGYGAGRDEGYWVISTQ 489
Qy 161 DGQVHLHGKSPQAR-----ISN-----PSQTTQTAQWLLLEASVSRG 197
Db 490 DGQRYFYGWGRSERTGATASVTFPEPVGNDTGTGCHDQFPEPCTQAWRSLDRAVDANE 549
Qy 198 EQIYYQYRAEDDTGCEADEITHLQATAQR-----YLIHIVYTG----- 235
Db 550 VETMYFY-----DKEYNHVRSVANSKAREYVSSGVYKEIQIGWSSQIPDGKLP 599
Qy 236 -----NRTASE-----TLPLDGSAPSQADWLFLYLVFDYGERSNL--KT- 273
Db 600 KVELSHVNRCTIRVOENDPLRDEPATCTPTDDKFTSPYDPEVVDLMCDGTSADYNCAKGY 659


```
Qy 254 ADWLFYLVDFYGBRSNNLKTTPAESTTTGSMLCRODRESRYEYGEPIRTRRLCRQVLMYHH 313
Db 722 A-----LTGNELO--PAY-----SIYHTKDAFK-----HT 744
Qy 314 LOALDSKITEHNGPTLVSRLL--ILNYDESA-----IASTLVFVRVVG-----HEODGNVVT 362
Db 745 L7DQDVNTN-----SGLGNLVNNEKANVVKADTKDMYKMLSTLDPLNNSDISVDK 797
Qy 363 LPPELAYQDFSP-----RHHAHWQ-----PMDVLANFNAIORWQ 397
Db 798 LDHINHLYLSLSAPDDTAIGNVRAMTNNELIRTSVWNPSPLEYTPPSNVLKTRPVLNPIK 857
Qy 398 LV-DLKGEGLPGLLYQ-DKGAWVRS-----AORLGEIGSD-----431
Db 858 VIKYKNSGRTRYIYEAPGKVMQSVNSITERSSITQNTPEITFDLYNRGTLPTGTYS 917
Qy 432 ---AVTWKMQPLSVIPSLQSNASLVNDINGDQLDWVITGFLRGYHSQRPDGSWTRFTP 488
Db 918 IRYATIWDENSEI--VRPTEQSLs-----HNNLELSYVIT--EDLSGN-----KKFVS 962
Qy 489 LNALPVYTHPRAQLADL-MGAGLSDL-----VLGPKSRLYXANTROGPAKGDVQ 540
Db 963 VIDVPFKIALAKAYASTLTIGKDAASFDKSDQVDNLG-ESVNLQTTAN--FTNSEGIILK 1020
Qy 541 SGGITLTP-----VPGADPRKLVAFSD-----VLGSGQAHLEVEYSATKVTCPN 583
Db 1021 ELIVTIPADNIKTNLTALIPDTEKRVVYTTDVRNGVNSNPTDLTKVTAVKYV-----1076
Qy 584 LGRGRFQOP-----ITLPGFSQPATEFNPQAQVYLAADLG-----SGPTDLIYV 626
Db 1077 -----FDEPLVLTNGQSFQTNMRVTVPEDAPILTKAH-SQIFTKGLDNTWLSGNKVELET 1130
Qy 627 HTNRLDI-----FLKSGNGPAPVTLRPFPEGLRFDHTCOLQADVOGLGVASLILSVPHMS 683
Db 1131 EDNRGDLVVKYTBESGNTIQNSLTSKGNKTEYN-----VSVPQM--1170
Qy 684 PHWRCDLTNNKPWLLNEMNNMGVHTLRRSSQFWLDBKAAALTTGOTPVCLPFP--741
Db 1171 -----IDRLN-----RHYKFRVDNQ--LDPTTHYAKGQTKIVNLIYE 1208
Qy 742 -----PIHTLMQT-----ETEDISGNK-----759
Db 1209 VFEGSVIADYKTTDGEVLSPLVTVNSQIEGTEYATPATIPDRVTFETTTDDGKVKKTIS 1268
Qy 760 --LVTTLRYAGAWDGRRERFRGFGYVEQDTHQLAQGNAPERTPPAL-----805
Db 1269 YHLISTPENOSGTVVGKQTIIEVHYVY-EPITTYEQIPNDAPQETPVALEVTYVDSEGNE 1327
Qy 806 -----TKNWATGLPVIDNALSTE-YWRDQAPAGFSPRT-----840
Db 1328 VOETEETHDAPGIIADKMQVTGQTAENGITTHVYQRIQSEIENAPQETPVALEVTYCY 1387
Qy 841 -----TWQDNKQVP-----LTPEDDNSRYWFRNALKGQLLSELYGLD 878
Db 1388 VDSBEGNEVQETEGTHDAPGIIIGDKMQVTGQTTEDGITHIYQK-----IQSE---IP 1438
Qy 879 DSTNKHVPYVTFER---SQVRRLQHTDSRYPLWSSWVESRNYHYERIASDPQCSQNIIT 935
Db 1439 NEAPQETPVALEVTYVDSBEGNEVQET-----EGTHQPPSIIGDKMQVTGQT 1486
Qy 936 LSSDRFQQLKQLSVQYPRROQPAINLPDTPDKLLANSY-DQOQQLRLTYQQSSWHH 994
Db 1487 TTADGI-----TYVVERIOSEIPNEAPKETPQLEVTYVDSBEGNEVQET---EETHH 1538
Qy 995 LTNNTVRLGLPDSRSDIFTYGAENVPAGGLNELLSDDKNLSIADDKPREYLGQ-QKTA 1053
Db 1539 APG-----IIG-----DKWYTGQTTESGITTHVYERIQSEIPNEAPQETPQLEVTY 1587
Qy 1054 YTDQNTTPTLOTP-----TFQALIAETITVFNQSTLSAENSIPSDKL 1097
Db 1588 YVNSBEGNEVQETEGTHQPPGIIIGDKMQVTGQTTADGITTYVYERIOSEIPNEAPKE-T 1646
Qy 1098 STTLEQAGYQQTNYLFPRTGDKVVAHGHGYDTGTAAQFWRPQKQSNQTLQTKITLIWD 1157
```

```
Db 1647 PVQLEVTYVDTD-----GNEVQETEGTHQPPGIIGDKW-----QYTGRTV-----1688
Qy 1158 ANYCVVVQTRDAAGLTTSAKYDWRFL-----TPVOL-----TDINDNQHLITLD 1201
Db 1689 -----EKDGITT---YVYERIQSAIPNEAPQETPQLEVTYVDTGNEVQETEE 1735
Qy 1202 ALGRPITL-----RFWG-----TENGKMTGYSPKASFPSPDVNNAALBELKKPLPVAQCV 1253
Db 1736 GTIQPRVIGDKWRYSGVTVTENG-ITKHVY-ERIQSKVPDA-----POETPQLEVTY 1789
Qy 1254 YAPESWMPVLQSQTFFNLAE-----ODWQKLYNARIITEDGRICTLAYRRVW 1300
Db 1790 VDPEG-----NEIQETTEGHQPPGIIGDRWQ--YTGKVTEDG-IITVYVER-I 1835
Qy 1301 QSKAIPOLISLNLNGRPLPHSLTTLTDRYDHDPEQOIRQOVVFSQGFGLLQAAARHE 1360
Db 1836 QSE--IP-----NNPQETPVELEVR-----YVDSBEGNEVQET---1867
Qy 1361 AGMARQORNEGSLIINQVHTENRWAVTGR-TEVDNKGQPIRTYQPYFLNDWRVYVNSDSAR 1419
Db 1868 ---TEGHQPPSII-----GDRMQYTGKVTEDG---ITTY--VYERIQSKVPNDAPR 1912
Qy 1420 QEKEAYADTHVDPFGIREIKVITAKGWFRRTLFTPWFTVNAEDENDMTAAEVKK 1471
Db 1913 VDIDELXITIYVDNNGREI-VPSRKGQLPPEQF-----IGDMQYTGKHIEK 1958

RESULT 15
US-10-156-761-7834
; Sequence 7834, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7834
; LENGTH: 2364
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-7834

Query Match 2.3%; Score 181.5; DB 4; Length 2364;
Best Local Similarity 16.8%; Pred. No. 0.00072;
Matches 274; Conservative 180; Mismatches 460; Indels 715; Gaps 77;

Qy 7 FSITELSLPKGGGAIQWGEALTPTGP-----DGMALSLPLPISA-GRGYAPAFITNY 59
Db 305 YAVTSSSSSDAGDY---RASTLSPTSGSWEVSTGSGSFNYSVP IQVPAPPMGSAFSLAMS 361
Qy 60 NS-----GAGN---SPFGLGWDNCVMTIRR---THGCVPH-----YDETDTFLGPE 100
Db 362 DSQSDGRTSASNNQASWSGMDNLVGFIERRYNRNCTDGLPTIGMCDSPNSAKEPD 421
Qy 101 GEVLVADQPRDESTLQINLGATFTVTGYSRSLRSHFSRLEYWQPKT-----TGKT-D 153
Db 422 GAAYVISLNGTSELIOQNNNGSAY-----HLKNDPGWRVQLPFDGHGAGRNGE 470
Qy 154 FWLIYSPDGQVHLIGKSPQARISN-----PSQTTQTAQWILLE 190
```

Db 471 YWISTQDQRYFYFGWGRSERTSTATASVFTVPVVVGNDAGBCHDQFPEPCTQAWRNLD 530
Qy 191 ASVSRGQIYYQYRAEDDTGCEADEITHLQATQRYLHVYVGNRTASETLPGLDGA 250
Db 531 RAVDANEVETWYFY-----DKEYNHYRSA----- 555
Qy 251 PSQADWLFLVFDYGERSNLKTTPAFSTGSLWLCRQDRFSEYGFPIRRLCRQVLM 310
Db 556 -----NTDKAREYVSSGYVKE-----IQGM----- 576
Qy 311 YHHLQALDSKITEHNGPTLVSRLLINYNDESALASTLVFRRVGHQDGNVVTLPLELAY 370
Db 577 -----SSQITDGVKPAKVELSHVN-----RCIERVOENDPLADEPAACP 616
Qy 371 QDPSRHHMQPMDVLAFNAIQRMQLVDLKEGLPGLLYODKGAWYRSAQRLGEIGS 430
Db 617 FDDKPGSYDP-VFVDLMDGTSAD-----YNACGKTYF----- 648
Qy 431 DAVTWEKQPLSVIPSLQSNASLVINDINGDGLDWDVITGFLRGYHSQRPDGSWTRPTPLN 490
Db 649 -----PTFFSTMDLMDI----- 660
Qy 491 ALPVEYTHPRAQLADLMGAGLSVLIGPKSVRLYANTRDGPAGKQDVVQSGDITLPPVG 550
Db 661 -----KTYVSDQDG-----TGWDLVQQYQNKYGMFN 686
Qy 551 ADPRKLVAFSDVLGSGOHLVEVSATKVTCPNLGRGRFGOPITLPGFSPATEFNPQAV 610
Db 687 PD-----GTIGKTLWL-----DYIQRT 704
Qy 611 YLADLDGSGPTDLI---YVHTNRLDIFLNKSGNGPAEPVTLRPPGELRFDHTCQLQMAVQ 668
Db 705 Y-----GDG-DDIVLVPVFNFRDLD-NKVS-----AELNPPR----- 736
Qy 669 GLGVASLILSVPHMSPHEHRCDLTWKCPWLLNEMNNMGVHHTLRYRSSQFPLDEKAAA 728
Db 737 -----IKEHGDIGATTVSYGFANACDIDLHPAQ 766
Qy 729 LTTGQTPVCYLFPPIHTLWQTEDEISGNKLVTLTRYARGAWDGREREPFGYVEQTD 788
Db 767 ASNTQD--CY-----WQWTP-----GETDSKTCWFKKF-LVTQV- 799
Qy 789 SHLQAGNAPERTPALTKNWTATGLPVIDNALSTE-----YWRDDQAFAGFSPRFTWQ 843
Db 800 -----QVDPVTATN--QDGAPEVMTTSYTYEDGAGWHFTNDPLIKDEDESWTDMR 846
Qy 844 DNKOVPLT-----PEDDNRVMPNRALKGOLLARSELVGLDDSTNKHVPVTVTEPRSQVRL 899
Db 847 GYQEVQVTTGAGAGOKTKKSWLYRLSG-----DRTSK----- 879
Qy 900 QHTDSRYPVLMSVSVESRNYHYERIASDPQCSQNTLSDDRPQOPKQLSVQYPRRQOPA 959
Db 880 -----ADASATKTVTV-DGDBG----- 895
Qy 960 INLYPDT--LPDKLLANSY-DDQORQLRLTYQSSWHHLTNVTVRVLGIPDSTRSDIFTY 1016
Db 896 -NNTYDSDLSGRIILSTLRDDTGTSHERTYHK-YWDH---NTAQYDGLPDAR-----FVR 946
Qy 1017 GAENYPAGLNLLELLSDKNLILADDKPREYLGQOKTAVTDGONTT---PLOTPT-ROALI 1072
Db 947 EKE-----TTNTKVVSSGW-REHTVETEDYDEGASITTFGLPMRTDDWGQSSV 993
Qy 1073 AFTETTVF-----NQSTLSAFNGSI FSKLS-----TTLEQAGYQOTNY----- 1111
Db 994 SDNRCTTYGRAVNTDNYDSTGAQRWTVLQDQVKHYSVGCSSIADSKQDGYASTLYDNATS 1053
Qy 1112 -----LFPRTGEDK-VWVAHGYTDYGTAAQFWRPQKQSNQTLTG-KITL 1154
Db 1054 VDANKPVDGNPTESRITYTKSGSYRSTW---SGYDDAGR---MWSDEGGSHNRTLTKYSPAN 1108
Qy 1155 IWDANYCVVQTRDAAG-----LTTSAKYDWRFL-TPVOLTDINDNQHLITLDALGRPI 1207
Db 1109 TWPLN-GVIVTTPDGPALPAHTALTSTAWTSRFWKGKPTSIQDANGNVTKMSLDAAGRLV 1167

Qy 1208 TLRFW-GTENGKMTGYSSPE-KASFSPSPSDVNAIELKKPLPVAQCQVYAPESMMPVLSQ 1265
Db 1168 EV--WRPTETG-----SSPSMKFSYTIPTNSA----- 1194
Qy 1266 KTFNRLAQDQWKLNYNARIITEDGRICITLAYRRVWQSKAIPQLISLLNNGPRLPPHSL- 1324
Db 1195 -----GVFDV-----DGYPHVATHVLO 1212
Qy 1325 TLTTDRYDHDPEQOIROQVWFSDGFGRLLOAAARHEAGM-ARQRNEDGSLIINVQHTNR 1383
Db 1213 SGTIVLSSH-----AYVDGLGRARETQTPMGNGVDAATGNEVNRQVSVTRYDSA 1262
Qy 1384 WAVTGRTE-YDNKG-----QPIRTYQPYFLNDW--RYVSN----- 1415
Db 1263 GNVGTSAVFRNQGTAGSGSPSAKVEDLPSYTDLVL-DWAGRAITSLQVNGASQDAGR 1321
Qy 1416 -----DSAROEKEAYADTHVVDPIGREIKVITAKGWFRRLTTPWFTVNEDE--- 1462
Db 1322 VDTTYDGDFTSVKNKVDAAADTYT-DVYGQVSKVVEHTG--SATYTTAYTTAKDELIKI 1378
Qy 1463 -----NDTA 1466
Db 1379 TDPRGNDTS 1387

Search completed: February 16, 2006, 21:54:17
Job time : 144.704 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:23:04 ; Search time 340.476 Seconds
(without alignments)
5259.201 Million cell updates/sec

Title: US-10-754-115-34
Perfect score: 13043
Sequence: 1 MYSTAVLNKISPTRDGQTM.....KALLESDIILHITYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	12396	95.0	2538	Q93RN7_XENNE	Q93rn7 xenorhabdus
2	12098	92.8	2524	Q8KSP0_XENNE	Q8ksp0 xenorhabdus
3	5486.5	42.1	2523	Q93RP0_XENNE	Q93rp0 xenorhabdus
4	5043	38.7	2516	Q9RN43_PHOLU	Q9rn43 photorhabdu
5	4986.5	38.2	2504	Q85160_PHOLU	O85160 photorhabdu
6	4914.5	37.7	2525	Q7N7Y9_PHOLL	Q7n7y9 photorhabdu
7	4802	36.8	2381	Q8GF92_PHOLL	Q8gf92 photorhabdu
8	4763.5	36.5	2378	Q7N7Y6_PHOLL	Q7n7y6 photorhabdu
9	4583.5	35.1	2376	Q9F923_9ENTR	Q9f923 serratia en
10	4572	35.1	2499	Q8GFA0_PHOLL	O8gfa0 photorhabdu
11	4515.5	34.6	2444	Q7N7Y1_PHOLL	Q7n7y1 photorhabdu
12	4491	34.4	2215	Q7N7Y0_PHOLL	Q7n7y0 photorhabdu
13	4061	31.1	2373	Q6XPA4_YERFR	Q6xpa4 yersinia fr
14	3948	30.3	2502	Q4ZNN7_PSESY	Q4znw7 pseudomonas
15	2965.5	22.7	2168	Q693A5_YEREN	Q693a5 yersinia en
16	1807.5	13.9	1197	Q8D1P7_YERPE	O8d1p7 yersinia pe
17	1779	13.6	1192	Q6QDZ3_YERPS	Q6qdz3 yersinia ps
18	1605.5	12.3	1362	Q7N493_PHOLL	Q7n493 photorhabdu
19	1595.5	12.2	1564	Q7MZV6_PHOLL	Q7mzv6 photorhabdu
20	1591	12.2	1565	Q85156_PHOLU	O85156 photorhabdu
21	1571	12.0	1189	Q85152_PHOLU	O85152 photorhabdu
22	1553.5	11.9	1615	Q87X47_PSESM	O87x47 pseudomonas
23	1507	11.6	1471	Q7N8C3_PHOLL	Q7n8c3 photorhabdu
24	1495.5	11.5	1617	Q4ZP57_PSESY	Q4zp57 pseudomonas
25	1332	10.2	1279	Q93RP4_XENNE	Q93rp4 xenorhabdus
26	781	6.0	2986	Q4HWU4_GIBZE	Q4hwu4 gibberella
27	758	5.8	147	Q8KSP4_XENNE	Q8ksp4 xenorhabdus
28	527	4.0	168	Q6WCC2_9ENTR	Q6wcc2 serratia pr
29	494.5	3.8	139	Q6WCC3_9ENTR	Q6wcc3 serratia pr
30	440.5	3.4	1095	O85151_PHOLU	O85151 photorhabdu
31	410	3.1	1080	Q66PW7_PHOLU	Q66pw7 photorhabdu

32 355.5 2.7 981 2 Q4ZP58_PSESY Q4zp58 pseudomonas
33 346 2.7 80 2 Q8KSP2_XENNE Q8ksp2 xenorhabdus
34 342.5 2.5 830 2 Q6QDZ2_YERPS Q6qdz2 yersinia ps
35 317 2.4 833 2 Q8ZAV3_YERPE Q8zav3 yersinia pe
36 309 2.4 774 2 Q8D1P8_YERPE Q8d1p8 yersinia pe
37 309 2.4 965 2 O85155_PHOLU O85155 photorhabdu
38 298 2.2 776 2 Q74PN4_YERPE Q74pn4 yersinia pe
39 292.5 2.2 966 2 Q7MPV5_PHOLL Q7mpv5 photorhabdu
40 292 2.2 1173 2 Q7N492_PHOLL Q7n492 photorhabdu
41 285.5 2.2 79 2 Q8KSP3_XENNE Q8ksp3 xenorhabdus
42 264.5 2.0 981 2 Q7N8C4_PHOLL Q7n8c4 photorhabdu
43 262.5 2.0 10203 2 Q5HPA2_STAEO Q5hpa2 staphylococ
44 261 2.0 4936 2 Q6D5C2_ERWCT Q6d5c2 erwinia car
45 254.5 2.0 4688 2 Q9PQ08_UREPA Q9pq08 ureaplasma

ALIGNMENTS

RESULT 1

ID Q93RN7_XENNE PRELIMINARY; PRT; 2538 AA.
AC Q93RN7_ XENNE
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE XptA2 protein.
GN Name=xptA2;
OS Xenorhabdus nematophilus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Xenorhabdus.
OX NCBI_TaxID=628;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX MEDLINE=21218513; PubMed=11319082;
RX DOI=10.1128/AEM.67.5.2062-2069.2001;
RA Morgan J.A.W.; Sergeant M., Ellis D., Ousley M., Jarrett P.;
RT "Sequence analysis of insecticidal genes from Xenorhabdus nematophilus
RT PMF1296.";
RL Appl. Environ. Microbiol. 67:2062-2069 (2001).
DR EMBL; AJ308438; CAC38404.1; -; Genomic DNA.
SQ SEQUENCE 2538 AA; 283995 MW; 3A7A98CF54FD756E CRC64;

Query Match 95.0%; Score 12396; DB 2; Length 2538;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2424; Conservative 31; Mismatches 71; Indels 24; Gaps 3;
Qy 1 MYSTAVLNKISPTRDGQTMTLADLYLSFSELKIPDDQLSMGEARHLHYHETIEQKNN 60
Db 1 MYSTAVLNKISPTRDGQTMTLADLYLSFSELKIPDDQLSMGEARHLHYHETIEQKNN 60
Qy 61 RLLEARIPTANPOLSGAIRLGIERSVSRSYDEMFGARSSSVFKPGSVASMFSPAGYLT 120
Db 61 RLLEARIPTANPOLSGAIRLGIERSVSRSYDEMFGARSSSVFKPGSVASMFSPAGYLT 120
Qy 121 ELYREAKDLHFFSSAYHLDNRRLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180
Db 121 ELYREAKDLHFFSSAYHLDNRRLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180
Qy 181 SDALMESLSTYRAIDTPYPHYPETIROVIMTHDSTLSALSRNPEVMQASGLAILA 240
Db 181 SDALMESLSTYRAIDTPYPHYPETIROVIMTHDSTLSALSRNPEVMQASGLAILA 240
Qy 241 NISPELVNLTTEETKKNADALFAQNSENITPENFASQSIKAYKYLGLSEVQKYLGM 300
Db 241 NISPELVNLTTEETKKNADALFAQNSENITPENFASQSIKAYKYLGLSEVQKYLGM 300
Qy 301 QNGYSDSTSAYVDNISTGLVNNESKLEAYKITRVKTDYDKNINYFDLMYEGNNOFFIR 360
Db 301 QNGYSDSTSAYVDNISTGLVNNESKLEAYKITRVKTDYDKNINYFDLMYEGNNOFFIR 360
Qy 361 ANFKVSRFEGATLRKNAGPSGIVGSLGPLANTNFKSNLYLSNDSYKNGKVKLYAVRY 420
Db 361 ANFKVSRFEGATLRKNAGPSGIVGSLGPLANTNFKSNLYLSNDSYKNGKVKLYAVRY 420

RA	Waterfield N.R., Daborn P.J., ffrench-Constant R.H.;	
RT	"Genomic islands in Photorhabdus.;"	
RL	Trends Microbiol. 10:541-545(2002).	
DR	EMBL; AF188483; AAF05542.1; -; Genomic DNA.	
DR	EMBL; AF346500; AAL18486.1; -; Genomic DNA.	
SQ	SEQUENCE 2516 AA; 282953 NW; 8250A0650B614B99 CRC64;	
	Query Match 38.7%; Score 5043; DB 2; Length 2516;	
	Best Local Similarity 42.6%; Pred. No. 3.7e-259;	
	Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;	
Qy	22 LADQYLSFSLRIRIFDQDLSWGARELYHETIEQKNNRLLLEARIETFRANPOLSGAIRL 81	
Db	21 LTDLSHSFNEFRQVSEHLSWSETHDYHDAQQAQKDNLYEARILKRAPOLQNAVHL 80	
Qy	82 GIERDSVSR-SYDMFGARSSSVKPGSVAMSPFAGYLYELYREAKDLHFSSAYHLDN 140	
Db	81 AILAPNAELIGYNNQFSGRASQYVAPGTVSMFSPAAYLTELYREARNLHASDVYLDLT 140	
Qy	141 RRPDLADLTLSQNMDEITSLTSLNELLLEHI-TRKTGSDSALMESLSTYRQAIDTPY 199	
Db	141 RRPDLKSMALSQQNMDELSTLSLSNELLLES IKTESKLENYTKMWMELSTFRPSGATPY 200	
Qy	200 HOPYETTRQVMTHTDSTLSALSRNPEVMGQAEGLAILAILANISPELYNIIITEIKNA 259	
Db	201 HDAYENREVLIQDQPLEQLNASPAIAGLMHQASLIGINASSPELFNIIITEITGNA 260	
Qy	260 DALPAQNFSENITPENFASQSWIAKYKYLELSEVQKYLGMQLN-GYSDSTSAVYDNISTG 318	
Db	261 EELYKKNFG-NIEPASLAMPKYLKRYNLSDEELSQFGRKASNFGQOEYSN---NQLITP 316	
Qy	319 LVVNNESKLEAYKTR-VKTDDYDKNINYPDLMYEGNQQPFIANFKVSRFEGAT----- 372	
Db	317 VNNSDGTGVKYRITREYTNAYQMDVELFP--FGGEN---YRLDYFKPNFYNASYLSIK 371	
Qy	373 -----LRKNAGPSGIVGSLSGPLIANTNPKSNVLSNISDEYKNGVKIYAYRTYSTS 425	
Db	372 LNDKRELVRTEGAQ-----VNIEYSANITLTAD--LSQPFELGLTRVLPSSG 418	
Qy	426 ATNQGGIFTFESYPLTIFALKNAIKRLCTLSGLSPNELQITVRSNDAQIINDSVLTK 485	
Db	419 WA-YAAAKFTVEEYNOYSFLLKNAIKRLSRATELSPTILEGIVRSVNLQDINTDVLGK 477	
Qy	486 VFYTLFVSHRYALSFDDAOVLGNSVINQYADDDSVSHENRLENTPLPKGKIFEADGTVS 545	
Db	478 VFLTKYMYQRYAIIHAETALILCNAPIQRSYDNPQSFQDRLENTPLNGQYFSGDEID 537	
Qy	546 IDPDEEQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVVISLSYRLTLIAR 605	
Db	538 LN-SGSTGDRWKTLIKRAFNIIDDVSLFRLKILTDHDKDGKIKNNKLSNLYIGKLLAD 596	
Qy	606 VHQTLVNELCMLYGLSPNGKTTAS-LSSGELPRVILWLYQVQWLTAEAITTBAILWC 664	
Db	597 IHQTLIDELDLLL-IAVGEKGTNLSAISDKQLATLIRKANTITSLWLTQKWSVFLPMT 655	
Qy	665 TPESGMSIPISNLIANNLRPSI---SEDMAQSHNRELQAEILAPFAATLHLASPDWAR 721	
Db	656 STSYNKTLTPEIKNLLTIVYHGLQFQDKKAD-----LLHVMAPYIAATLQLSSENVAH 709	
Qy	722 YILLANTNLRPGGLDIAGFMT-----LVLESINANETQ-----LVQFCHVMAQLSLS 770	
Db	710 SVLLWADKLQPD-----GAMTAEFWDWLNKYTPGSSAEVETQEHIVQYQALAQLEMV 765	
Qy	771 VQTLRLSEAEISLVIVSGFAVLGAKNPAGQHNIDTLFSLYRFHOWINGLNGNPGSDTDM 830	
Db	766 YHSTGINENAPFLVTKP-EMFGAATGAAPAHDAISLIMLTRFADWVNALGEKASSVLAA 824	
Qy	831 LRQOTLTADRLASVMGLDISWVTQMVSA-----GVNQLQCODINTVLOWIDVA 880	
Db	825 FEANSLTAEQADAMNDANDANLLQASIQAOHQHLPVPTPENAFSCWTSINTILLQWNVAA 884	
Qy	881 SALTMTPSVIRTLVNIKRVVTAALNAESNLPDSEWQTLAENNEAGLSLQQOQTLLADYTA 940	

Db	885 QQLNVAPOQGVSAVLGLDYIQSMKET-----PTYAQWENAAAGVLTAGLNSQQANTLHAFIDE 940	
Qy	941 RLSSVLNCWFLANIQPEGVSLHREDDLYSYFLIDNOYSSAIAKTTRELAETAGIOLYNRA 1000	
Db	941 SRSAALSTYYIROVAKAAAAIKGRDDIYQYLLIDNQVSSAAIKTRIAEATASIOLYNRA 1000	
Qy	1001 LNRIEPNARADVSTROFFTDW-TVNNRYSTWGGVSRVLYVYPENYIDPTQIGQTRMMDEL 1059	
Db	1001 LENVEENANGSVLSRQFFIDWDKYNKYSTWAGVSQVLYVYPENYIDPTMRIGQTKMDAL 1060	
Qy	1060 LENISQSKLSDRTVEDAFKTYLFRFETVADLVVSAVHDNNSNTGTLTFWVGQTRNLPE 1119	
Db	1061 LQSVSQSLNADTVEDAFMSYLTSEFQVANLKVISAHDNINNDQGLTYFGLSETDAGE 1120	
Qy	1120 YYWENVDI SRMQAGELAAANKWKTKIDTAVNPVKDAIRPVIPEERLHLIIVWEKEEVAK- 1178	
Db	1121 YYWRSVDHSKPNQDGFPAANAWSEMHHKIDCPINPYKSTIRPVIYKSRILYLLWLEQKTKQ 1180	
Qy	1179 --NCTD--PVETDYDRFTLKLAFLRHGDSWSPMSYDITQVEAVTDDKPDPTERLALAASG 1234	
Db	1181 TGNKSGQYQYETDYRYELKLAHRYDGTWNTPTTFDYNKKISEL--KLEKNRAPGLYCAG 1238	
Qy	1235 FQGEDTLLVYFYTKGYSDFGSGSNKNVAGMTIYGDGSFKKMENTALSRYSQLKNFTDII 1294	
Db	1239 YQGEDTLLVMFYNQDITLDSY--KNASMQGLYIFADMASKDMTPEQSNVYRD--NSYQQF 1294	
Qy	1295 HTQGNDLVRKASVRFADQFVFPASLANGS--ATGDDSLTWENGNIPQITSKYSSDNLAI 1352	
Db	1295 DTNN---VRVNNRYAEDYEIPSSVSRKDYGMWDYLSVMYNGDIDPTINYKAASSDLKI 1351	
Qy	1353 -----TLHNAFTVRYDGSNNVIRNKQISAMKLTGVGDKSQYGNAFIANTVKHYGGY 1405	
Db	1352 YISPKLRIIHNG-----YEQG-----KRNQCNLMNKYG-----KLGDKFIV-----Y 1388	
Qy	1406 SDLG-GPITVYNKTKNY-IASVQGHLMNADYTRRLILTPVENNNYAKLFEP-PPSPNTIL 1462	
Db	1389 TSLGVNPNNSGNKLMFPVYQYSGNTSGLNQR--LLFHRDITTPSKVKEAMIPGAKRSLT 1446	
Qy	1463 NTVFTVGS-----NKTSDFKCSYAVDGNNSQGFQIFSSYOSSGWLDDITGNNNDIK 1515	
Db	1447 NQNAAGDDYATDNLNPKDPLKQYIFMTD---SKG-----TATDVSGGVEINTALSPAKVQ 1499	
Qy	1516 ITVMAGSKTHTFTASDHASLIPANSFDAMPYTFKPLEIDASSLAFTNNIAPLDIVFETKA 1575	
Db	1500 IIVKAGKEQFTFADKDVSIQSPSPDEMYQFNALIEDGSLNFNNSSASIDVTFATAFA 1559	
Qy	1576 KQGRVLG-----KIKQTLKSVKRVNVPEDIIFLRETHSGAQVMQGVTRIRINTLLASQLV 1631	
Db	1560 EDGRKLGYESFSIPVTLKV-----STDNALTLAHNENGAQYMQWQSYRTRTLTLFAQLV 1614	
Qy	1632 SRANTGIDTILTWETQRLPEPPLGEGFFANFVLPKYPABEHGDERWFKIHIGNVGGNTGR 1691	
Db	1615 ARATGTIDTILSMETQNIQBPQLGKGYATFVIPPYNLSTHGDERWFKLYIKHVVDNNS- 1673	
Qy	1692 QPYVSGMLSDTSETSMTLF-----VPYAEGYMHGVRVLGVGYQKITDYNTWESAFFYD 1746	
Db	1674 HIIYSGQLTDTN-INITLFIPLDDVPLNQDYH----AKVYMTFKKSPSDGTWGWGPHFVRD 1728	
Qy	1747 ETQOQFVLINDADHDSGMTQOGIVKNIKKYKGLNVSIAIATGYAPMPFNSSALYIYELF 1806	
Db	1729 D--KGIVTIN-----PKSILTHPESVNLNNIS-----SEPMDFSGANSLYFWELF 1772	
Qy	1807 YITPMCFORLLQKQDEATQWNYVNPAGYLVNGEIAPIWNCPCPLETSTTWNANPL 1866	
Db	1773 YITPMLVAQRLHEQNFDEANRWLKYWSPSGYIVHGQIQNYQWNVAPLLEDTSWNSDPL 1832	
Qy	1867 DAIDPDAVAQNDPMHYKIATFMRLLDQILIRGDMAYRELTDRDALNEAKWYVTRTLELGD 1926	
Db	1833 DSDVPDAVAQNDPMHYKVSVTFMRLDILLIARGDHAYQLERDTEINEAKWYMQALHLLGD 1892	
Qy	1927 EPEDYGSQQAAPSLSGAASQTVQAAQODITMLGRG---GVSKNLTANSILVGLFLPEY 1983	
Db	1893 KPYPPLSLTTSWSDPRLDRAADITTTQNAHDSAIVALRQNIPTPAPLSLRSANTLTDLFLPQI 1952	

```
QY 1984 NPALTDYQTLRLKRLFNLRHNLSDGQPLSLAIYAETDPKALLTSMVQASQGGSAVLPG 2043
D 1953 NEVMNMYQTLAQRVYNLRHNLSDGQPLSLAIYAETDPKALLTSMVQASQGGSAVLPG 2012
QY 2044 TSLYRFPVMLERTRNLVAQLTOFTSLLSWAHHDADDELITLLLOQGHMELATQSIROQ 2103
D 2013 FMSLWRFPMLLENARGMVSQTLTQFTSLTIQIERQDAEALNALLQNAELILTNLSIQD 2072
QY 2104 RTVDEVDADIALLAESRSANRLKRYOQLYDDEINHGEQRAMSLLDAAAGSLAGQVLS 2163
D 2073 KTIIELDKTVLEKSKAGASRPDSYCKLVNENAGENQANTLRASAGLITAVQASR 2132
QY 2164 IAEGVADLVPNVFLGACGSRWGAALRASAVMSLSATASQVSADKISRSEAYRRRRQEW 2223
D 2133 LAGAAADLVPNIFGAGGSRWGAALAEATGYVMEPSANVMNTEADKISQSEYRRRRQEW 2192
QY 2224 EIORDNAGEVQMDAQLESKIRREAAQOMQVEYQETOQAHTQAQLELLQKFTNKALYS 2283
D 2193 EIQRNABEELKQIDAQLKSLAVREAAVLQKTSLKTOOBTQSQAELQKFSNQAALYN 2252
QY 2284 WMRGKLSAIYQFDFLTOSFCMLAQEALRRELTDNGVTFIRGGAWNGTTAGLMAGETLL 2343
D 2253 WLRGLAIIYQFYDLAVARCLMAEQAYRWELNDDARSFIPKGAQGYTAGLLAGETLL 2312
QY 2344 NLAEMEKWMLRDERALEVTRTVSLAQFYQALSSDN--FNLTETKLTQFLREGKNGVAGS 2401
D 2313 SLAQMEDAHLKRDKEALREVERTSLAEVYAGLPKONGPFSQAQEDKLVLSQSGSAGSN 2372
QY 2402 NELKL-----SNRQIEASVRLSDKIPSDYPESIGNTRQLQKQVSVTLPALVGPVEDIRAV 2456
D 2373 NNLAFGAGTDTKTSIQASVSFADLKIREDPASLGKIRRIKQISVTLPALGYPQDVQAI 2432
QY 2457 LNYGSSIVMPRGCSAIALSHGVNDSPQFMDPNDSRYLPPFGISVNDSSGSLTLSPFDDAT- 2515
D 2433 LSYGDKAGLNGCEALAHSHGVNDSSGQFLDPNGKFLPFFEGIAL-DOGTTLTLSPFNASM 2491
QY 2516 ---DRQKALLESLSDIILHRYTIR 2537
D 2492 PEKGQATMLKTLNDIILHRYTIK 2516
RESULT 5
O85160 PHOLU PRELIMINARY; PRT; 2504 AA.
AC O85160
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-FEB-2005 (TrEMBLrel. 29, Last annotation update)
DE Insecticidal toxin complex protein TcBa (Toxin complex protein).
GN Name=tcbA;
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photobacterium.
OX NCBI_TaxID=29488;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RX MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;
RA Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bhartiya R., ffrench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photobacterium luminescens."
RL Science 280:2129-2132(1998).
RN (2)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W-14;
RA Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,
RA ffrench-Constant R.H.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN (3)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
```

```
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrench-Constant R.H.;
RT "The tc genes of Photobacterium: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF047457; AAC38627.1; -; Genomic DNA.
DR EMBL; AF346438; AAL18460.1; -; Genomic DNA.
DR InterPro; IPR002345; Lipocalin.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
SQ SEQUENCE 2504 AA; 829654 MW; 8F29B1693D1047CE CRC64;
Query Match 38.2%; Score 4986.5; DB 2; Length 2504;
Best Local Similarity 42.3%; Pred. No. 3.7e-256;
Matches 110; Conservative 434; Mismatches 863; Indels 217; Gaps 48;
QY 9 NKISPTBDG-----QTTWTLADLQYLSFSELRKIFDDQLSWGEARHLHYETIOKKNRL 63
D 3 NSLSTIDTICQKLQLTCPAIEALYPPDTPREKTRGMVNWGEAKRIEYIAQAEQDRNLH 62
QY 64 EARITRANPOLSGAIRLGIERD--SVRSYDEMFGARSSFFKPGSVASMFSPAGLYTEL 122
D 63 EKRIFAVANPLLNKAVRLGTQMLGFTIQGYSDLFENRADNYAAPGVSASFSPAAVLT 122
QY 123 YREAKDLHFSSAYHLNRRPDLADLTLSQSNMDEITSLTSLNELLEHITRTKGTGSD 182
D 123 YREAKNLHDSIIYLDKRRPDLASLMSQKMDDEISTLSNELCLAGIETKTGSD 182
QY 183 ALMESLSTYRQAITPYHQYETIRQVIMTHDSTLSALSRNPVMOGAEGASLLAILANI 242
D 183 EVMDLSTYRLSGETPYHAYETVREIYVHRDPGRHLNQAPIVAALDPTVLIGISSHI 242
QY 243 SPELYNIITEBITKN---ADALFAQNFSENIPTENFASQSWIAKYIYGLSEVOKYLG 299
D 243 SPELYNLIIEIPKDEAALDTLYKTNFGD--ITTAQLMSPSYLARYYGVSPEDIAVYVTS 301
QY 300 LQN--GYSDSTSAVDNISTGLVNVNESKLEAYKIRTKVTDDYDKNINFDLMYEGNNOFF 358
D 302 LSHVGYIS-----SDLIIVPLVDGVGKWEVVRTRTPSDNITSTQTNIELYPPQGDNYL 354
QY 359 IRANFKVSRFEGAT---LRKNAGPSGIVGSLSGP---LIANTFNKSNLYNISDSYKNG 412
D 355 IKYN--LSNSFLDGFYLYQKDGSDWTEIAHNPVPMVINOQYESQATIKRSDS--NI 410
QY 413 VKIYAYTSTSTATNOGGGIFTFESYPLTITFALKLKAIRKCLTSLGSLPNELOTVRSD 472
D 411 LSLGLQRWHS--GYSNFAAANFKIDQSPKAFLLKMKAKAIRLKATGSLFATLERIVDSV 468
QY 473 NAOGIINDSVLTQVYTLFYSHRYALSFPDDAQVINGSVINOYADDSDVSHENRFLNTPPL 532
D 469 NSTKSIITVEVLNRYVYKFIYDRIGISEETAAIILANINISQAQVNGQLSQEQLFNHPL 528
QY 533 KGKIFE-ADGNT-----VSIDPEEQSTPARGALMRGLGVNSGELYQL-----GKLAG 579
D 529 NGIRVEISEDNSKHLNPNDLNPKDSTGDDQKAVLKBAFQVNASELYQMLLITDRKEDG 588
QY 580 VL--DAQNTITLTVFVSSLYRLTLARVHOLTWNELCMYGLSPFNKGTATSSSGSLP 637
D 589 VIKNNLEN-----LSDLYLVSLAQIHLNLTAEINLILVICGYGDTNITQITDDNLA 640
QY 638 RLVIWLVQVQWLTAEAITTEAIVLLCTPESFGNISPEISNL-----LNNLRPSISED 690
D 641 KIVTLLWITQWLTQKWTVDLMTATYVTLTPEISNLATSLTSLHKGESLIGED 700
QY 691 MAQSHNRELOAEILAPFATATHLASPDMAFYILLWTDNLRPGGLDIAGFMVLVKESLN 750
D 701 LKRA-----MAPCFTSALHLTSQSVAYDVLWIDQIQPAQITVDGFW-----EVQ 746
QY 751 ANETT-QLVQPCNMAQLSLSVQTLRLSEALSVLVSGFVAVLGAQKQACQAGNIDITLFS 809
D 747 TTPTSKLVITPAQLAQLSLIYRRIGLSELSLIVTQSSILLVAGKS--ILDHGLLTMA 804
QY 810 LYREHONGINGLGNPGSDTLDMLRQQTLTADRLASVMGLDIDSMVTOA--MVSAGVNOLOCW 867
D 805 LEGPHTWVNGLGQHASLILAAKDKGALTVDVAQAMNKEESLLQMAANQVKEKDLTKLTSW 864
```


DR	InterPro; IPR007087; Znf_C2H2.	QY	940	ERLSSVLCNFWFLANIQPEGVSLHSDDLYSYFLIDNQVSSAIKTTTLAABAIAIGIYYINR	999
DR	PROSITE; PS00697; DNA_LIGASE_A1; UNKNOWN 1.	QY	940	ERSAALSTVYIRQVAKPAAAIKSRDLYQYLIDNQVSSAAIKTTTAAIAIASIQLYVNR	999
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN 1.	QY	1000	ALNRTEPNARADSVTRQFETDW--TVNNRYSTWGGUSRLVYVYENYDPTQRIQGTQMTMDE	1058
KW	Complete proteome.	Db	1000	ALNVEENADSGVISRQFIDWDKYNKRYSTWAGVSQLVYYPENYDPTIRIGQTKMDA	1059
QY	SEQUENCE 2525 AA; 285096 MW; 892D6877B7FE3B0 CRC64;	QY	1059	LEENISQSXLSDTVDADAFKTYLTREPETVADLKVVSAYHDNVNNSNTGLTWFGVQGTRENLP	1118
QY	Query Match 37.7%; Score 4914.5; DB 2; Length 2525;	Db	1060	LQGSVQSQQLNTDTTVEDAFWSYLTSEQVANLKVISAYHDNINNDQGLYFYFGLNETDAG	1119
QY	Best Local Similarity 41.3%; Pred. No. 2.6e-252;	QY	1119	EYVWRNVDISRMQAGELAAANAKEMTKTIDTA VNPYKDAIRPVIFRERLHLIWEKESEVAK	1178
QY	Matches 1091; Conservative 461; Mismatches 829; Indels 259; Gaps 58;	Db	1120	EYVWRVSDHSKFSKSDGFAANAWSEWHKIDCPINPYRSTIRPVYKSLYLLMLEQKEITK	1179
QY		QY	1179	---NGTD--PVETDYDRFTLKLAFRLHSDGWSAPWSYDIT---TOVEAVDTKKDPDTERLAL	1230
QY		Db	1180	QTGNSKDGQYQTDYRYELKLAHRYDGTWNTPIITFDVNEKISELSEKLVNKPAP-----GL	1234
QY		QY	1231	AASGFGQEDTLLAFV-----YKTGKS-----YSDFGSGNKNVAGMTIYDGSFQKKN	1276
QY		Db	1235	YAGYQGEDTLLVFMFYNQODTLDYSYKTAAMQGLYIFADWASKDMTLGQSNVYRDNYSYKQF	1294
QY		QY	1277	ENTALSRYSQLKNTFDIIHTQGNDLVRKASRYRPAQDFEVPASLNMGSAL--GDSDSTVME	1334
QY		Db	1295	DTNS-----VRRVNNRYAENYEIPSSVNSNNSYGRGEYNLSMVH	1333
QY		QY	1335	NGNIPOITSKYSSDNLAITLHNAAFVRYDGGNGVIRNKQISAMKLTGVDGKSKQYGNFI	1394
QY		Db	1334	GGDPIIRYQASSDLKHI-SPKLRITVEGADERTRQ-----	1371
QY		QY	1395	IANTVKHYGYSDLGGPIVYNTKONYIASVQGHLMNADYT--RRLILTPV-----ENN	1446
QY		Db	1372	---CNLLKYG---KLGDKFIYTNLS-----VNTQWTSNKKUIYPYQYRGNTTNN	1417
QY		QY	1447	YVARLFEF-----PFSPN-----TILNTVFTVG--SNKTSDFKCKSYAVDGN	1487
QY		Db	1418	SQGRLLFYRDNQIGKVSFLPWGQSLYSEVNSNNGYIIDTSNQEPFLKQLYMEDKNN	1477
QY		QY	1488	SGQFQIFSS--YQSSGLDIDTGINNTDIKITVMAGSKTHTTASDHIAASLSPANSFAMP	1545
QY		Db	1478	--GFRIATDVYNA--VEINTAISPAKIQVTWAGSKETKTADKDSVQSPSPSPDEMY	1532
QY		QY	1546	YTFKPLEIDASSIAFTNNIAPDIIVPETKADGRVLGKIKQTLVKRVNYPNEDILPLRE	1605
QY		Db	1533	YQFNTLEBEGTGLNFNNNLASIDVFTTALAERDKLGYESFSIPVTR--KVSTDNALTLHH	1591
QY		QY	1606	THSGAQYMQLGYYVIRLNTLLASQLVSRANTGIDITLTMETQRLPEPL--GEGFFANFVL	1664
QY		Db	1592	NENGAQYMQWQAYRFLNTLFAQLVARATGIDITLSMETQNIQEPVLWSDNFTA--TL	1649
QY		QY	1665	PKYDPAEHGDERFWKXIHIGNVGNGTRQP-----YSGMLSDTSETSMTLFVPAEYGYM	1719
QY		Db	1650	PIYNSSLHGNKNIKML-----GNSEKQPDQANVIYQGBLTQKNQL-----	1691
QY		QY	1720	HGVRILGVYQKI-----TYDNWESAFY--FDETKQFVLINDADHDSGHTQOQIV	1770
QY		Db	1692	---INLSETYIISKERSNNTKPDVNDKNSDWLYLVKYSTNPDKIGEFWFKNGTNPQGY	1748
QY		QY	1771	---KNIKKYKGLNVSATGYSAPMDFNASALAYWELFYVTPMCMFORLQOEKDFDQATOW	1829
QY		Db	1749	PKGNENRD--PIFIENKQTEPMDPFGSANTLFWELFYTPMVLVAQRLHQBQNFDEANRW	1806
QY		QY	1830	INYYNPGAYIYNGEIAPIWNCRPLEETTSWNPALDAIDPDVAQNDDPMHYKIATFMR	1889
QY		Db	1807	LKYVWSPSGYI VHGIQYQNMVNRPLLEDTSNNSDPLSDVDPDAVAQHDPMHYKVSTFMR	1866
QY		QY	1890	LDDQLILRGDMAYRELTRDALNEAKWYVRLTELLGDEBEDYGSQWAAAPSLSGAASQTV	1949
QY		Db	1867	TLDDLITARGDQAYRQLERDRLNEAKWYQNALHLLGDKFYLPLSTTNNPRLDRAADITT	1926

Db 892 DWDKYNKYSTWAGVSQLVYYPENYIDPTMRIGOTKMDDTLQSVSQSLNADTVEDAPK 951
 Qy 1079 TYLTFETVADIKVVSAYHDNNSNTGLTFWQOTRENLPEYWRNVDSISMAQAGELAN 1138
 Db 952 SYLTSFEQVANLEVISAYHDNNDNDQGLTFYFGNSKTEVQYWRSDVHSKFDGKPAAN 1011
 Qy 1139 AWKEWKIDTAVNPYKDAIRPVI FRERLHLIWEKEBAKNGTDPVETD - RFTLKLAFL 1197
 Db 1012 AWSEMHKIDCAINPYQSTIRPVIYKSLYLIWLEQKETAKQEDNKVTTDYHYELKLAHI 1071
 Qy 1198 RHDSWSAPWSYDITTOVEAVTDKPDTERLALASFGQEDTLVAVYVTKGKSYDFGG 1257
 Db 1072 RYDGTWNPITFDVDEKILALELTK - SOAPGLY CAGYQGEDTLILIMFYRKKEKLDY - 1127
 Qy 1258 SNKNVAGMTIYDGGFKKMENTALGRYSQKNTFDIIHTQGNLVRKASYRPAQDFEYPA 1317
 Db 1128 KTAPOGFVIFSDMGSKDWTNEQCNRYRNGYTHFDTNSDINSVIR - INNRYAEDYEIPS 1186
 Qy 1318 SLNMGSA - IGDSLTVMENGNIPOITSKYSDNLAITLHNAAFVTVRVDGSGNVIRNKQI 1375
 Db 1187 LINSNSHDWGEYNSQVYGGNI - VINYKVTNSDLKIYI - 1224
 Qy 1376 SAMKLTGV - DGKSQGNAPFIANTVKHYGG - YSDLG - - - - - GPITVY 1415
 Db 1225 - SPKLRI IHGKE - - GRERIOSNLIKYKGLGDKFIIYTSLGINPNNSNRPMFYPVYQY 1281
 Qy 1416 NKTNYIASVOCHLMNADYTRRLILTPVENNYARLFBF - PPSNPNTILNTVTVGSG - - - - 1470
 Db 1282 NGNTSGLA - QGRLL - - - FHR - - - - - DTSVSSKVAWI PCAGRLSINENANIGDDCAE 1329
 Qy 1471 --- NKTSPFKCSYAVDGNNSOGFOIFSSYSGSGWLDIDTGINNTDIIKITWAGSKTHFF 1527
 Db 1330 DSVNKPDDLKQYIYMTD - - - SKG - - - - - TATDVSGPVDINTAISSEKVOITIKAG - KEYSL 1381
 Qy 1528 TASDHASLPANSPAMPYTFKPLEIDASSLAFTNNAIPLDIVETKAKDGRVLG - KIKQ 1586
 Db 1382 TANKDVSQVSPFSFEMCYQFNALSIDGNSLNTNNSASIDVTFALADDGRKLGYEIPN 1441
 Qy 1587 TLSVKRVNVPEDILFLRETHSGAOYMGOLGVYRIRLNTLLASQLSVSRANTGIDTILMET 1646
 Db 1442 IPVIOKV - KYDNALTLPHDENGAYQMWGAYRIRLNTLFRQLYERANTGIDTILNET 1499
 Qy 1647 QRLPEPPLEGFFANFVLPKYDPAEGBDERWPKHI GNVGNGTQOPYYSGMLSTSETS 1706
 Db 1500 QNIQPMWGIAGYIELIDKYNPDTHGNTGSKFIYIGDIFRAGDHFPIYQGALSIDTQT 1559
 Qy 1707 MTLFVPYAEYVME - GYRLGVGYOKIYDNTWESAPFYFDETKOQFVLINDADHSGMT 1765
 Db 1560 VKLFLPRVDNAYGNKNLYVYAAYQKV - - - - - ETNFIRFVKEDNNKPKAT 1603
 Qy 1766 QQGIKNIKKYKGFNLVSIATGYSAPMDFNSASALYYWELFYTPMCMCFQRLLOEKQFDE 1825
 Db 1604 FDTYKN - GTPGLASARVIOVSEPMDFSGNSLYFWELFYTPMVAQRLHQNDFE 1662
 Qy 1826 ATQWYNNYPAGYVINGEIPWIKNCRPLETTSNNANPLDAIDPDAVAQNDPMHYKIA 1885
 Db 1663 ANRWLKYVWSPSGYVIRGQIKNYHNVRPLLENTSWNSDPLDSVDPDAVAQNDPMHYKVA 1722
 Qy 1886 TFMRLLDQILRGDMAYRELFEDALNEAKMYVTRLELLGDEPEDYGSQWAAPSLSGNA 1945
 Db 1723 TFMRLDLMARGDHAYRQLERDTLINEAKMYWQALHLLGNKPYLPPLSSVWMDPLDAAA 1782
 Qy 1946 SQTQAAVQOQDLMGRGVSQKN - - - - - LRANSVLVLFLPEYNPALDYQWTLRLRLNLRH 2003
 Db 1783 ATTTQKHAAYITSRQGTQTPALLIRGANTLTLFLQINDVMSYNWKLRLNLRH 1842
 Qy 2004 NLSIDQPLSLAIYAEPTDPKALITSMVOASQGSASVLPGLTSLYRFPVMLERTNLVAQ 2063
 Db 1843 NLSIDQPLHLPIYATPADPKALLSAAVATSQGGKLPESFISLWRFPHMLNARSMTQ 1902
 Qy 2064 LTQFTSLSWAHDADDELITLLLOQOMELATQIRIQORTVDRVDADIABLRSSA 2123
 Db 1903 LTQFGSTLQNIITEROAESLALLQNAKELITLTLISIQDKTIEIDAETKTVLEKSAGA 1962

Qy 2124 QNRLEKYOQLYDEEDINHGEORAMSLDDAAAGQSAGVQ - -LSIABGVADLVNPFGLACG 2181
 Db 1963 KSRFDNYSKLYDEDVWAGERQALDM - -RIASQITSGLKGLHMAAAALEMVPNIYGFVAG 2020
 Qy 2182 GSRWGAALRASAVMSLSATASQYSDAKISSEAYRRROEWEIQORDNADGEVKQMDAOL 2241
 Db 2021 GTRYGAIANATAIOGGIAAEGLLIEAEKVSOSEIWRRRRQBEWEIORNNAEAMKQIDAOL 2080
 Qy 2242 ESLKIRRAAQAQVQYEQYQTAQHTQAQLELQRFKTKALYSMWRGKLSAIYQYFDFDTQ 2301
 Db 2081 KSLTVREAAVLQKTGLTKTQEQEOTAQLAFQKESNOALYNWLRGLAAIYFQYVLDLV 2140
 Qy 2302 SFCMAQALRELTNDNGVTFIRGAWNGTTAGLMAGETLLILNLAEMKVMLEDERALE 2361
 Db 2141 ARCLMAEQAYRWETNDSARPFKGAWQGTYAGLAGETLMLNLAQMEDAHLKQBRAL 2200
 Qy 2362 VTRVTSIAQFOALSSDNPNLTKLQFLREGK - GNVGASGNELKLSNRQTEASVRLSDL 2420
 Db 2201 VERTVSLAAQVYSGEKSFAUKKIEALLQGDKETSGNDGNQLKJTNNTSATUTLQDL 2260
 Qy 2421 KIFSDYPE - -SIGNTRQLKQVSVTLPALVPYVEDIRAVLNTYGGSIIV - MPRGCSAIALSHG 2477
 Db 2261 KLKDDYPEMQLGKTRIKQISVSI.PALLGPYQVQAVLSYGGDATGLAKGKALAVSHG 2320
 Qy 2478 VNDGQFMLDPNDVRYLPFEGISVNDGSLTSLSPDPDATROKALLESLSIILHRYTI 2536
 Db 2321 LDNGQFOLDPNDGKFLPFEGIDINDKGTFTLSPFAAASKQKNIILQMLTDIILHRYTI 2379

RESULT 8

Q7N7Y6 PHOLL
 ID Q7N7Y6 PHOLL PRELIMINARY; PRT; 2378 AA.
 AC Q7N7Y6;
 DT 01-MAR-2004 (Tremblrel. 26, Created)
 DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
 DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
 DE Insecticidal toxin complex protein TcdA4.
 GN Name=tcdA4; OrderedLocusNames=plu0965;
 OS Photorhabdus luminescens (subsp. laumondii).
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Photorhabdus.
 OX NCBI_TaxID=141679;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=TT01;
 EX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
 RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
 RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
 RA Dassa E., Derosé R., Derzelle S., Freyssinet G., Gaudriault S.,
 RA Médigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
 RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
 RT luminescens";
 RL Nat. Biotechnol. 21:1307-1313 (2003).
 DR EMBL; BX571862; CAE13260.1; -; Genomic_DNA.
 DR Photolista; plu0965; -;
 KW Complete proteome.
 SQ SEQUENCE 2378 AA; 271176 MW; 09834DDDBF1D4C6B CRC64;

Query Match 36.5%; Score 4763.5; DB 2; Length 2378;
 Best Local Similarity 40.5%; Pred. NO. 2.7e-244;
 Matches 1047; Conservative 447; Mismatches 795; Indels 293; Gaps 47;

Qy 22 LADLOYSFSELKTFDDOLSGEARHLYHETIEOKKNRLLLEARIFTRANPOLSGARL 81
 Db 21 LTDCHVSFNRQVQVSDHLSWSENRLYRDAQQEQENQLYEASILKRANPOLQNAHL 80
 Qy 82 GIERDSVS - RSYDEMFGARSSSFVKPGSVASNFSPAGYLTLYEAKDLHFSSSYHLDN 140
 Db 81 GITLPHAELOQYNEFGGASQYVAPGSVSMFSPAALTYLYEARNLHASDSVYHLDQ 140
 Qy 141 RRPDLADTLTQSNDMTIEISTLTLSNELLLEHI - TRTKTGGSDA - LMESLSLYROAIDTP 198

```

141 RRPDLQSWLNOQNDTELSTLSLSNELLGCVKQNGPNDLNDTKVWAMLSFPSPSGTIP 200
199 YHQPYETIROVIMTHDSTLSALSNPVPMQAGASILAILANISPELYNLTBEITEKN 258
201 YHDAYENVRKAIQODPQKQFQSPAVAGLHQASILGINNISPELFINLTBEITEAN 260
259 ADALPAQNFSENITPENPASOSWAKYGL---ELSB-VQKYLGMQLONGYSDSTSAYDN 314
261 AEDVYKKNFGD-IDPTWLAPEYLKSYNLSDELSQFIRY----- 301
315 ISTGLVNVNKSLEAYKITRVTDDYDKNINYFDMLYMEGNQPFIRANFKVSRFEGATLR 374
302 -----PNNE----- 305
375 KNAGPSGIVGSLGSLPIANTNFKSNLYLSNISDSBYKNGVKIYAYRYTSTSATNOGGGIF 434
306 -----LNTQ----- 309
435 TPESYPLUTIPALKUNKAIRCLTSGLSPELQTTIVRSNDAQIINDSVLTKVFTLYPSH 494
310 -----KIHLKINKIKILLSRSVNLFPVKLDEIPEQS-----INPSVLGKIFLVKYMQ 358
495 RYALSFDQAQVLNGSVLNOYADDOSVSHENRFLNTPPLKGKIPEADGNTVSIIDPDE-EQS 553
359 KYNIDTETALILCNNSISQSYRNQPSQFDRFNTPPLNGOCFTVEDTNDLNLNSTDT 418
554 TFARSALMRGLGVNSGELYQLQGLAGVLDAQNTITLSVFVSIISLYRLTLARHQLTVNE 613
419 NWHKTVLKRAFNVDISLYRLHLHANHENKDKIANNIKNLSNLYMTKLLADIHQLTIDE 478
614 LCMYGLSPFNGKTTASLSGSELPRILWILYQVOWLTEABITTEATLWLLCTPFPSGNIS 673
479 LYLLMTITGEDKRDLYTIDDKLEKLNRDLSLSNMLRTQKWSYQLFMTTNYDKTLT 538
674 PEISNLLNLRPSISEDMAQSHNELQAEILAPFAATLHLASPDMAHYILLWDLNLRPG 733
539 PEIQNLLNTVYNGLON---PDENKDLLTISPPIAATLQLPSENVASHILLWADKIKPS 595
734 GLDIA-----GFMTLVLKESLNAVNETTQLVQFCHVMAQLSLSVOTLRLEABLS 782
596 EKKITTEKFDWLQNRG-----ILELSKPPEMOEQIIQVCHCLAQLTWIRSYGINENAFR 651
783 VLVSIGFAVLGAKNQ---ACQHNIDTLFSYRHOWINGLGNPGSOTLMDLROQTITAD 839
652 -LFIEQPTIFGISDEPKATPAHNAPTLIILTFEANWVNTLGEKASPLTAFENKILTA 710
840 RLASVMGLDLSMVTQAMVSA-----GVNQLQCQWDINTVLOWIDVASALHTMPSVIRT 892
711 QLANMMLDANLLSQASIQAQNYQVTKENTFWSWQSIDIILQWNTNVANDENISFQGIS 770
893 LVNIYYVTALNKAESNLPMSDEWQTLAENMBAGLSLTOQAOTLADYTABRSLVLCNWFLA 952
771 LIALDYIQPSQKT---PTYSQWENAAWVLTAGLDTQQMTLHAFDLBSRSTALSNIYIS 826
953 NIQEGVSLHSRDDLYSYFLIDNOVSSAIKTRLAEBAGIAGOLYINRANLRIENARADV 1012
827 KIANRAAAIKSRDDLYQYLLIDNOVSSAIKTRIAEASIASIQLYVNRALENIETHAISDV 886
1013 STROFFTDW-TVNNKSYTWGVSRLVYYPENYIDPTORIGOTRWMDLELLENISQSKLSRD 1071
887 ITRQFFDNDKYNKRYSTWAGVSQLAYPENYIDPTWRIGOTKMDTLLOSVSQSOLNAD 946
1072 TVEDAFKTYLTRFETVADLKVVSAYHDVNSNTGLTWPFVGOTRENLPYIYWRNVYDISRMQ 1131
947 TVEDAFKSYLTSFEQVANLEVISAYHDVNVNDQGLTFVIGHSKTEINQYWRNSVDHSGFN 1006
1132 AGELEAANAKWETKIDTAVNPKDAIRPVI FRELHLIHWKESEVAKNGTD-PVETYDRP 1190
1007 DGKFAANAWSWHKIDCAINPYQSTIRPVIYKSRLLYLWLQSKIAQDKDKNKVITDIYY 1066
1191 TLKLAFLRHGSGWAPWSYDITTOVEAVTDKPKPOTERLALAASGFQGEDTLTLLVYVYTKG 1250

```

```

1067 ELKLAHIRYDGTWNTPIAFDVNSKISALELAK--SOAPGLYCAGYQGEDTILLVMFYRKKE 1124
1251 SYSDFGSGNKNVAGMTIYDGSPKQMENTALSRYSQLKNTFDIIHTOGNDLVRKASVRPA 1310
1125 KLDDY--KTAPMOGFIYFSDMSKEMTNEQNSYRNGYTHFDNTNSDTSVIR-INNRYA 1181
1311 QDPFVPASLANMGA--IGDDSLTWENGINIPOITSKYSSDNLAITLHNAAPVRYDGSNG 1368
1182 EDYEIPSLNHSNDHGEYNLSQVYCGNI-----VINEYVASS 1220
1369 VIRNKQISAMKLTGVGDKSQYGNAPIIANTVKGHGYSDLGGPITVY-NKTKNYIASVOG 1427
1221 DLKIKISPKLRIIH-EGKE--GRERIQCDLMKYG---KLGDKFIIYTNINKNPNHKSQE 1274
1428 HLMNADYTRRLILTPVENNYARLFE-----PPSPNTILNTVTVGSKNTSDF 1476
1275 KLIYVPVQYR-----NNDKGLLFYRSNTTGTIRAFWPTTKBEITAT-----TSSN 1320
1477 KKSUYAVDGNSSQGFQIFSSV-----QSSGWLDDIDTGINNTDIKITVMAGSKT 1524
1321 QDC--IIDTINN--IDLKAYFYMEDKEGQIATDISATDINTGISSEKVOIINAG-KE 1375
1525 HTFTASDHIASLPANSPDAMPYTFKPLEIDASSLAFTNNIAPLDIVFETKAKGRVLG-K 1583
1376 YSLTANKDVSQVSPSPSEEMCYQFNALIEDGSSLNFNNGASIDVTFIALADDCRKLGE 1435
1584 IKQTLVKRVNYNPEDILPLRETHSGAQMQLGVYRIRLANTLLASQLSVRANTGIDTILT 1643
1436 IFNIPVIOKY--KTDNALTLFHDKNQAQYQWGAAYRVLNLTLFARQLVERATTGIDTILS 1493
1644 METQRLPEPLGEGFFANFVLPKYDPAEHGDERWFKIHGNVGGNTGRQPYSGMLSDTS 1703
1494 METONIQEPMMGIGAYIEI VMDKXNTDIGHNTKNSFKIWMYGDIFKAGDHPIYQOTLSOIT 1553
1704 ETSMTLFPVYABGYMHE-GVRLGVGYQKITYDNTWESAFYFDETKQQFVLINADAHDS 1762
1554 QTTVKLPLRVDNAYGNKNLNVYAAVOKA-----ETHFRFKEDNNKNPATDPTTNN 1607
1763 GMTQOGIVKIKKYKGPLNYSIATGYSAPMDPNSASALYVWELFYPTPMCFORLLQEQ 1822
1608 G-----TPFGLASARVITQTFSEPMDFGANSLYFWELFYPTMPLVAQRLLEHQN 1656
1823 FDEATOWINYVNPAGYIVNGETAPWNCRLPETTSWNPANPLDAIDPDAVAQNDPMHY 1882
1657 FDETNRKLVKWSPSGIYVGOIKYVRNVRPPLFENTSWNSDPLDSVDPDAVAQNDPMHY 1716
1883 KIATFMRLQLLQILRGDMAYRELTRDALNEAKMYVTRLELLGDEPEDYGSQQWAAPSLS 1942
1717 KVATFMTLDDLMMARGDHAYRQLERDNLNEAKMYMQALHLLGDKPYLPLSSAWNPRLD 1776
1943 GAASQTVQAAVQODLTWMLGRGVSKN--LRTANSLVGLFLPEYNPALTDYMQTLRLRFPN 2000
1777 NAADTTIQKAYSAITLRQGTQTPALLRSANTLTDLPQIINDVMLSYNKKLELRLYN 1836
2001 LRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASQGSASVLPGLTSLRYFPFVMLERTNL 2060
1837 LRHNLSDGQPLYLPMYATPADPKALLSAVATVSQGGSLPQPFWSLWHFPHMLENARM 1896
2061 VAQTQFTGTSLLSMAEHDDADELTTLTLLQOQMEIATOSIRIQRTVDEVDADIABESR 2120
1897 VAQTQFTGTSLLTIERQDAEALNALLQNAKELILTSLTQDKTIEIDAETVLEKSK 1956
2121 RSAQRLEKYQOQLYDEIDINHEGEQBSMLSDAAAGSLAGOV--LSIAGEVADLVNPFGL 2178
1957 AGAQRSDFNYSKLYDEVDNAGERQALDM--RIASQTSITSLGKLUHMAAALEWPNYIGF 2014
2179 ACGRSRGAALRAGASVMSLSATASQYSADKISSEAYRRRRRQEWIEIQORDNADGEVKOMD 2238
2015 AVGSTRFGALANATAIGGIAAEGLLIEAKVSQSEIWRERQEWIEIQORNAEAMKQID 2074
2239 AOLESKIRREAAQMOVEYQETOQAHTQAQLELLQRTKALYSWRGKLSAIYQFFD 2298
2075 AQOQSLTVRREAAVLQKTSLKTOEQOTQAQALAFQKRFESQALYNLWLRGLAAIYFQFYD 2134

```

Matches	1056;	Conservative	411;	Mismatches	810;	Indels	355;	Gaps	55
Qy	1	MYSTAVLLNKISPTR-----DGO	TMTADQLVYLSFSELRKIFDDQLS	WGEGARHLYHETI	54				
Db	6	MYNIDDLILEKVNAPEARLSENDTAVTLTDLFSRSFPVKKITCDLSLMGCEVCYLVSQAQ	65						
Qy	55	EQKNWRLLEARIPTANPOLSGAIRGLTERDSVSRSDYMFGAGSSSFVKPQSVASMPFS	114						
Db	66	HEOKENRLETESRLABANPLLVNAVRLGRQAAGSRSYDDWFGRADRPARPGSVASMPFS	125						
Qy	115	PAGVITELVREAKDLHFSSAYHLDNRRPDLDLTLSQSNNMDETEISTLTLSNELLEHHIT	174						
Db	126	PAAYITELVREAKDLHPDTSLFRDLIRRPDLAALALSQNMMDELSTLSLSNELLYRGIG	185						
Qy	175	RKTGGSDALMESLSTYRQAIPTYHQPYETIRQIVIMTHDSTLSALSARNPEVMQARGAS	234						
Db	186	AASEGLDDDSVRELLAGRYLTGLTPYHWAEARQAAILVDQPTLMGFGRNPDPVAQLMDPAS	245						
Qy	235	LIALIANISPELYNYILTEBITEKNADALPAONFSENITPENFASQSWIAKYVGLELSEVQ	294						
Db	246	MLAIEADISPELYQILAEBEITIDSYEALMSKNFGD-MPPSSLASYDALATFYDLDYDELTA	304						
Qy	295	KYLGMLQNGYSdstSAVDNISGLVNNEsk--LEAYKI	TRVKTDDYDKNI	NVFDLMWE	352				
Db	305	SLLS-LRLDFSNPNNEYINSQLSVTNLNESTGLTI	IHHYRLTGd-SQINP-ELIPY	361					
Qy	353	GNNQFFIRANFKVSRFBGATLRKNACPGSIGVGS-----LSGPLIANTFNKSNYLSNIDS	407						
Db	362	GDGTily--NFsvv----STISEDSPKLGSLSGNSSNLYSGDYQLQGVrYsIPVEIDEG	415						
Qy	408	EYKNGWKIVAYRVTSSTSATNQGGI-----FTFESYPLTI	FALKNKATRLCLT	SGLSP	462				
Db	416	KLMDGITI-----GLSRKGGGYSTWNFTLI	EYDPAIFILKLNKVI	RLRYKATGMTT	466				
Qy	463	NELQTIVRSDNAOGIINDSVLTKVFYTLFVSHRYVALSFDDAQVLNGSVINOYADDSDVS	H	522					
Db	467	ABIYQITNIIINGLTIHDHAVLSKIFLVRYLMRHVQLDVARSLLICNGTISDOAFSGBTGL	526						
Qy	523	FNRLFNTPLPKXGIHEADGNTVIDPDEOESTFARSALMRGLGVNSGELYQLQGLAGLVLD	582						
Db	527	FTTLFNTPLNGQLFSADDTPLDRSEAPEDAFLSVLKKEAFNISASGLSTLWLQASG-D	585						
Qy	583	AQNTITLSVFEVSISSLRYRLTLRARVOLTVNELCWLGLSPFNKTKTASLSGSELPRUVIW	642						
Db	586	SSAGFSCSADNAAALRYVKLLADIHDLsAGELSMLLSVPSPFGVAAGSLSDNELTQ--F	642						
Qy	643	LYQVTOWLHEABITTEAWILLCTPFPSGNISPIISNILLANNLRPSISMEDMAOSHNRLEOA	702						
Db	643	LYQTITLWTEQGTVSDVFMLMTQTQGLTLPDIENLLASLRNGLS-----GRELFPE	695						
Qy	703	IL---APPTAAATHLASPDMAVYILLWTDLNLRPGGLDIAGFWTLVLKESINANETTQLV	758						
Db	696	TLPGDGAPPFAAAWQIQDADTAKAMLTWADQLAPEGLUTTEFILVVNNAAPNDEQACOMA	755						
Qy	759	QFCVMAQSLSVQTLRLSBAELSVLVISGFVLAGKNQPAG-----QHNIIDLTFSL	810						
Db	756	GFCQALWQALAIIRSTGLSTRFLTIV-----SQPGRFRTGWHHLPHDLPALRDI	805						
Qy	811	YRFHOWINGLNGPESDTLMLROOTLTADRLASVMGLDISMWTOAMV--SAGVNO--L	864						
Db	806	TRFHVNVNRSCHAGEVLTALETGELSALLARALSQNEODVTGALAQRVAGREQDQNSVF	865						
Qy	865	QCWODINTVLOWIDVASAHMTESPVRTLVNIRVVTALNKAESNLPSWDESMOTLAENMEA	924						
Db	866	TSWEEVDQAEOQLMDNETSITPSGLASLTALKY---NVSDOSAPLYSQNWVVSGLLQA	922						
Qy	925	GLSTQQAQITLADYTABRSLSVLCNWFIANTQPSGVLSHRSDDLLYSYFLIDNQVSSAIKTT	984						
Db	923	GLAKSQSSALHDYLEECTSSALCAYILRNLANPMVS--GRDDLFGYLLLDNQVS	980						
Qy	985	RLAETAIGIOLYNRALNRJEPNARADVSTQFTDW-TVNNRYSTWGGVSRVLVYPYENY	1043						
Db	981	RIAEATAGIRLYINRALNGIELSAMAEVRQOFETDWDTPNKGYSTYWAGVELSYVYPYENY	1040						

RESULT 9
 ID Q9F923_9ENTR PRELIMINARY; PRT; 2376 AA.
 AC Q9F923;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE SepA.
 GN Name=sepA;
 OS Serratia entomophila.
 OG Plasmid PADAP.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Serratia.
 OX NCBI_TaxID=42906;
 RN [1]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ALM02;
 RX MEDLINE=20416224; PubMed=10960097;
 RX DOI=10.1128/JB.182.18.5127-5138.2000;
 RA Hurst M.R., Glare T.R., Jackson T.A., Ronson C.W.;
 RT "plasmid-located pathogenicity determinants of *Serratia entomophila*,
 RT the causal agent of amber disease of grass grub, show similarity to
 RT the insecticidal toxins of *Photobacterium luminescens*.";
 RL J. Bacteriol. 182:5127-5138 (2000).
 RN [2]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ALM02;
 RX MEDLINE=22960473; PubMed=14597010; DOI=10.1016/S0147-619X(03)00062-3
 RA Hurst M.R., O'Callaghan M., Glare T.R.;
 RT "Peripheral sequences of the *Serratia entomophila* PADAP virulence-
 RT associated region.";
 RL Plasmid 50:213-229 (2003).
 RN [3]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ALM02;
 RX PubMed=15262948; DOI=10.1128/JB.186.15.5116-5128.2004;
 RA Hurst M.R., Glare T.R., Jackson T.A.;
 RT "Cloning *Serratia entomophila* antifeeding genes--a putative defective
 RT prophage active against the grass grub *Costelytra zealandica*.";
 RL J. Bacteriol. 186:5116-5128 (2004).
 RN [4]
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=ALM02;
 RA Hurst M.R.H.;
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF135182; AAG09642.1; -; Genomic_DNA.
 KW Plasmid.
 SQ SEQUENCE 2376 AA; 262648 MW; E418DACE22DBB7BF CRC64;
 Query Match 35.1%; Score 4583.5; DB 2; Length 2376;
 Best Local Similarity 40.1%; Pred. No. 1.1e-234;

```
Qy 1044 IDPTORIGOTRMMDELLENISQSLSDTVDKFTYLTRETVADLKVVSAVHDNVNSN 1103
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1041 LDPTVRIGQTMMDTLQSVSQSINRDTVEDAFKTYLTTFEQIANLNTVSGYHDNASMT 1100
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1104 TGLTFVFGQTRNPEYFYRWVDISRMQAGSLAANAKWKTKIDTAVNPYKDAIRPVIFR 1163
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1101 QGTYWYGRSITDQTNWYRSANHSKIODSMMPANAWGTWKINGMNPWSDLVCSVFFN 1160
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1164 BRHLHIWKEVEKANGNDPVETVDR-FTLKLAFLRHGDSWSAPWSVDITTOEAVTDKK 1222
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1161 SRLVYVWVEENQSDATEASTTTTQQSYTLKLSFRYDGTWSSPSVFDITGNIAF----- 1215
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1223 PDTERLALAAAGFOCEDTLLFVYVTKGYSYDFG-----SNKNVAGMTIYGD-GSEKKME 1277
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1216 PETQGMVTCNPLT-EQLYCAFYVTSK--PDFDNAQLISVDNDMTLNVISDGIKFSVS 1272
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1278 ---NTALSRYQLKNTF-----DIHTQGNDLVRKASYPAPQDFEVPAS 1318
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1273 HEFNTISTEKF--INNVSFSDPSANYFVSATSLLDDVIHS-----DFSL--- 1312
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1319 LMGSAIGDDSLTWENGINIPOITSKYSSDNLAITLHNAATVTVYDGSNVRNKQISAM 1378
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1313 LN-----SKTSTVFTNEDSSLLTPE-----LHITANVSCFVSTAGIATOSTIEKF 1358
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1379 KLTGVD-----GKSQYG-NAFI--IANTVKHYGYSYDLGGPITVYNTKKN-IASV 1425
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1359 VOAGIEFEINFYAQAAGGDFGVGDVNSKVTQVQKEAVG--VTV--KSYVTGV 1412
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1426 QGHLMNADYTRRLILTPVENNYARLFEFPFSPNTILMTVTVGSKNTSDFKKCSYAVDG 1485
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1413 SGSVE-----LFIDSNKYFSGIL-----SDKMITALISGST-----SKVNY-VSS 1452
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1486 NNSQGFQIFSS-----YQSSGMLDIDTGINNTDIIKTVNAGSKTHPTFASDHIALSPA 1538
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1453 IGSQDFSVKSLMPALQIYELIDILLTSGVNGTEIK-----SWPSAEWYNDKLSLQSG 1506
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1539 NSFDPAMPYTFKPLEIDASSLAFTNNIAPL-----DIVPETKADGR-VLGKIKQTLGVK 1591
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1507 NNL-----FNTKSLSFVTNTSDIVDEDEVDVTFVTFAVDQNNVLAARTAILTVI 1555
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1592 RVNYPNEDILFLRETHSQAQWQL-----GVYRIRLNTLLASQLSRANTGIDTILTMETQ 1647
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1556 RNINNDTSVIALRKNTRCAQIIRFTAGNDVALIRLNTLFAQLVDRANTGIDTILSMETQ 1615
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1648 RLPEPPLGEGFANFVLPKPYDAEHGDERWPKIHI GNVGNTGRQPYSGMSLSTSETSM 1707
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1616 RLTEPALFEEG-----SDVF----- 1629
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1708 TLFVPYABGYMHGVRILGVGYQKITDNTWESAFFYFDETKQFVLINDADHDSGMTQQ 1767
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1630 ----- 1629
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1768 GIVNKIKYKGLNVIATGYSAPMDFNSALYWFELYYTPMCFQRLQEQFDEAT 1827
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1630 -----MDPSGANALYFWELYYTPMVFQRLLEQHFPEAT 1665
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1828 QWINVYVNPAGYIYNGETAPWINCRPLEETTSNNANPLDAIDPDVAQNDPMYKATF 1887
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1666 RWLQTVNPNPAGVNVGLQNTWNVRLPEEDTGWNDSPDLSIDPDATAQYDPMYKATF 1725
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1888 MRLLDOLLRGDMAVRELTRDALNEAKMWYVYRLELIGDEPEDYGSQQAAPSLSGAASQ 1947
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1726 MSYDLILLIARGDAAYRLRLDERTLNEARWYVQALMLLGDPEYISFDADWSALITGDARASE 1785
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1948 TVQAAAYQODLWMLGRGGVSKNLRTANSVLGFLPEYNPALPDYQTLRLRLFNLRHNLISI 2007
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1786 VTRRDYQEAALLAVRFLVPAPEPTRTANSITLFLPQQNEVLKGYQTLQAQLHNLNRHNLISI 1845
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2008 DQOPLSLIAEPTDPKALLTSMVQASOGGSANVLPGLTSLRFPVWLERNLVAQLTQF 2067
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1846 DQOPLSLSVYATPSPSLQSAVNSQAQAALPAAVMPLYSFPVMLENARGMVSLTGF 1905
```

```
Qy 2068 GTSLLSMAEHDDADDTLLTLQQGMELATQSIHQRTQTVDEVDADIIVLAESRKSQAQNL 2127
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1906 GNTLLGITERQDAEALAKLIQTQGSSELIHQGLQODNVLSIEDADIAALEBSRGAQMRP 1965
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2128 EKYQOLYDEIDNHGEQAMSLLDAAAGSLAQGVLSIAEGVADLVNPNVFLGACGSRWGA 2187
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1966 ERYKVLVEADVNTGEKQAMDLYLSSSVLSASTAALFLAEAAADMLPNYIYGLAVGGSRYGA 2025
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2188 ALRASAVMSLSATASQYSADKISRSEAYRRRQEWIORDNADGEVKQMDAQLESKIR 2247
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2026 LFKATAGIQVSSDATHISADKISQSEYVRRRREWEIORDSAQSDVAQIDAQAAWAVR 2085
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2248 REAAQMVQYQETQQAHTQAQLELLQKFTNKALYSWMRGKLSAIYYQFFDLTQSFCLMA 2307
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2086 REGAELQKTYLETOQTOQAQALFLOSKFNNTALYSWLRGRLSAIYYQFYDLVAVSRCLMA 2145
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2308 QEARURRLTNGVTFPIRGAWNGTTAGIMAGETLLILNLAEMEKVWLERDERALEVTRTVS 2367
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2146 QQAWQMDKFETR-SFIQFGAWMGANAGLAGETLMLNLAQMEQAWLTGDERAIEVTRVC 2204
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2368 LAQFYQALSSD-NFNLTETKLTQFLREGKGNVAGSGNELKLSNRQIEASVRLSDLKIFSDY 2426
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2205 LSEVYTSIAEDAAEASLADKVVVELVNSGSGAGTKSNGLQMDQQOLEATLKLADIGINDY 2264
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2427 PESLGNTRQLKQSVTLTPALVGPYEDIRAVNLVGGSIWMPRGCSAIALSHGVNDSGQFML 2486
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2265 PVSILGTWRRRIKQISVTLTPALVGPYQDVRAVLVSGSVMPRGCSALAVSHGMNDSGQFOL 2324
Qy :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2487 DFNDPRYLPPEGLPVDVTGTLTSLFPDADGKQAMLLSLSDIILHIRYTTIS 2538
Db :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2325 DFNDPRYLPPEGLPVDVTGTLTSLFPDADGKQAMLLSLSDIILHIRYTTIS 2376

RESULT 10
Q8GFAO PHOLU
ID Q8GFAO PHOLU PRELIMINARY; PRT; 2499 AA.
AC Q8GFAO;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE TcdA2.
GN Name=tcdA2;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=23488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA french-Constant R.H.;
RL "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RL "Genomic islands in Photorhabdus.";
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AAQ17201.1; -; Genomic_DNA.
DR InterPro; IPR000577; FGSY_kin.
DR InterPro; IPR001063; Ribosomal_L22.
DR PROSITE; PS00445; FGSY_KINASES_2; UNKNOWN 1.
DR PROSITE; PS00464; RIBOSOMAL_L22; UNKNOWN 1.
SQ SEQUENCE 2499 AA; 283745 MW; E7C4C920146E4D8 CRC64;

Query Match 35.1%; Score 4572; DB 2; Length 2499;
Best Local Similarity 38.7%; Pred. No. 4.8e-234;
Matches 1063; Conservative 444; Mismatches 740; Indels 502; Gaps 72;
```

Qy 22 LADIQYLSFSELKRFDDQLSWGBARHLYHETIEBQKNNRLLEARI FTRANPQLSGAIRL 81

Db 890 IAEASTQLYINRALKNMEEDTDSVTSRSTFTNWDKYNKRYSTWASIAKLLYPENVI 949
 QY 1045 DPTORIGOTRMDELLENISQSLSRDVEDAFKTYLRFETVADLVKVVSAHYHDVNSNT 1104
 Db 950 EPTLRIQGTQMDALLQSISQSLQNTDVEDAFKSYLTSFEQVANLEVISAYHDNINNDQ 1009
 QY 1105 GLTWVFGQTRNLPYVYRNVDISRMQAGELAAWKEWKIDTAVNPYKDAIRPVIPIRE 1164
 Db 1010 GLTYIGRSKTEVQYVRSVDHNFSEKGFANAMSEWHKIDCPINPYEDTIRPVIQOS 1069
 QY 1165 RLHLWVWEKEVA-----KNGTDPVEYDRFTLLKFLRHDGWS 1204
 Db 1070 RUYITWLEOKVTHRSSEGTLPKVALEIKTDKQSKTTSY-YELKLAHRIADATWN 1128
 QY 1205 APWSYDITTVQBAVDKPKDPTER-LALASGFGQEDTLIVFYKTKGYSDFGSGKNVA 1263
 Db 1129 TPIITFDV--KKISDLNLEANKVLGLYCASYQGRKULLVMFYQKPEQASY--INAPMQ 1183
 QY 1264 GMTIYGDGSKOM--ENTALSRYSQLKNFTDIIHQGNLDLVRKASYRPAQDFEVPASLNM 1321
 Db 1184 GLYISSNMSQEBMNPNDYKLNHKOLDT-----HTATNSVIR-VNNRYAESYEIPSSVNN 1237
 QY 1322 GSAI--GDSLTVMGNGIPIQTSKYSSDNLAI-----TLHNAAFVTVYDGSNGVLRN 1372
 Db 1238 NNGYDWGEGYLSWYVGGSI--LITRNQSDNSKIQISPKLRIIHNG-----YEGR----QR 1286
 QY 1373 KQISAMKLTGVGDKSQYG---NAFTIANTVHYGYSDIGSPITVYN-KTRNYIASVOGH 1428
 Db 1287 NOCNLMK-----KYGPKDKFII-----YTLG--INQNNLSNKKFIYPVQY 1327
 QY 1429 LMAADYTR--RLILPVENNYVARFEPFPPSPNTILNTVTVGSKNTSDFKKCSYAVDGN 1486
 Db 1328 EGNASQSGRLFLYRDSATNFWRAW-LPNLPSGQEMSITVGGNISGNYGVIDNKHNSN 1386
 QY 1487 NSQGFQIFSSY-----QSSGLDIDFGINNTDIIKIVMAGSK-----THFTTAS 1530
 Db 1387 NP-----FKGYFYMDHGGIDTVDVSGPMPINTNOPSNNKIITVTKVEDGKLTSEPYTAN 1441
 QY 1531 DHIASLPANSPAMPYTKPL-EIDASSIAFTNFIAPIDIVPETRAKDGRLVGLKIKQLIS 1589
 Db 1442 DKVSVKPLNFEMCYQFNGLEIDVSKLVFNKNEASIDITP-TAFADAPEGQKERNLG 1500
 QY 1590 -----VKRVYNPEDILFURETHSQAQYVQWGLVYIRLNTLTLAASQLVSPANTGIDTI 1641
 Db 1501 EBHFSIRIUKKSETN--NVLSLHNSSGAQYQWQAYRTRNLTPARKLISRAAGIDTI 1558
 QY 1642 LTWETQRLPEPLGLGFFANFVLPKYDPAHEGDEWFKHIGNVGNTGRQRYYSGLMSLD 1701
 Db 1559 LSMETQNTQELQLGKGFYATFVIPPYNPSTHGDWRWFKLYIKHVSDDNS-HIYISGQLKD 1617
 QY 1702 TSETGNTLFVP-----YAEGYMHGVRGLGVGYKITYDNTWESAFYFDETQKQ 1751
 Db 1618 TN-IGITFIPLDDIPLNQDFHAKY-----MTFKQSPSDGSGWSPHIRD--NKG 1665
 QY 1752 FVLINDADHSGMTQOGIVKNIKKYKGFNLVNSIATGYSAPODFNSASALYWEFYFPTM 1811
 Db 1666 IITINSS-----SILTHPESVNVNLNVG-----SEPMDFNGANSLSYFWEFYFPTM 1711
 QY 1812 MCFQRLQKQDEATOWINYNYPAGYVNGEIA-PWITNCRPLEETTSWNAANPLDAID 1870
 Db 1712 LIAQRLLEHQNFDEANRLKLYVNFSGHIANQKQYPHITWVVRPLOEDTSWNDDPLGSPD 1771
 QY 1871 PDAVQNDPMHYKIAFTFRLLDQLILRGDMARELTRDALNEAKWYVTRLELLGDEPED 1930
 Db 1772 PDAIAQYDPMHYKVASFRTLDLLEQGYAYRQLERTDLAEAKWYQALHLGDKPHL 1831
 QY 1931 YGSQQWAAAPSLSGAASQTVQAAVQDQLTWLGRGVSKNLRTRANSVLGLFLPYNPALTDY 1990
 Db 1832 SLSTWNPDELEEAADLEKQTHAKAVADLRQ-----QSTGGSNTDLFLPQVNWMLSY 1886
 QY 1991 WOTLRLFLNLRHNLSDIQPLSLAIYABPTDPKALLTSMWQASQGSVNLPTGLSLRFP 2050
 Db 1887 WOKLQRLYNLRHNLSDIQPLHLPIFATPADPKALLSAVASSQGGNNLPSAFISVWRF 1946

QY 2051 PVMLERTNLVAQLTQFGTSLLSMAEHDDADELTTLLLOQGMELATQSIQRIQORTVDEVD 2110
 Db 1947 PHMLENARSMVSQLTQFGSTLQNIETTERQDAEALNTLLOQAELVLTNLSIQDKTIEED 2006
 QY 2111 ADIAVLASRRSAONLEKYQOYLVDEDINHGEORAWSLDDAAAGOSLAGOVLSIAGVAD 2170
 Db 2007 VEKTVLEKTRAGARFDSYSKFYNEDINAGEKQAWLRASVAGISTALQASHLAGAALD 2066
 QY 2171 LVPNVFGLACGSRGAALRASASVMSLSATASQYSADKISRSEAYRRRRQSEIQRDNA 2230
 Db 2067 LAPNIFGADGGSHWGAQAATGNVWEPFASVMNTEADKISQSEAYRRRRQSEIQRDNA 2126
 QY 2231 DGEVQMDAQLESKIRREAAQMVQYQETOQAHTQAQLELLQRFKTNKALYSWRGKLS 2290
 Db 2127 ETELQIDAQLGSLVVRREAAVLQKTSKLTQEQTHAQLTFLQHKFSNOALYNWLRGLS 2186
 QY 2291 AIYVQFPDLTQSFCLMAQOEALRRELTDNGVTPIRGAGNNGTTAGLMAGETLLNLAEEMK 2350
 Db 2187 AIYVQFQDLYTVARCLMAEMAYRWETNDAAARFIKPGAMOGTHAGLAGETLLNLQAQMD 2246
 QY 2351 VWLDERALEVTRTVSLAQFYQALSSDNFNLTETKLTQFLREGKGNVGASGNELKL---- 2406
 Db 2247 AHLKQEQRVLEVERIVSLANIYK--EKQFSLTEKIAELMKNKLVTAGSGNNTLKFGTGN 2304
 QY 2407 SNRQTEASVRLSDKIFSDYPES--LGNTRQLKQVSVTLPALVGYEDIRAVLNYGGSIV 2464
 Db 2305 AGTSLQASISLADLQIRHDPENSGVGNVRIKQISVTLPALVGYQVQVQAILSYSGDVT 2364
 QY 2465 -MPCSCSAIALSHGVNDSGQMLDENSDRYLPFEGISVNDGSLTSPEDA-----TDR 2517
 Db 2365 GLAESCKLSVSHGVNDGQFQDFNDGKFLPFEIGIDI-DKGTLLTSLFPNALGEGSKGK 2423
 QY 2518 OKALLESIDILHIRYTIIR 2537
 Db 2424 QXTMLESIDILHIRYTIIR 2443

RESULT 12

Q7N7Y0 PHOLL PRELIMINARY; PRF: 2215 AA.
 AC Q7N7Y0
 DT 01-MAR-2004 (TREMELrel. 26, Created)
 DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
 DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
 DE Insecticidal toxin complex protein TcdA5.
 GN Name=tcdA5; OrderedLocusNames=plu0971;
 OS Photobacterium luminescens (subsp. laumondii);
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Photobacterium.
 OX NCBI TaxID=141679;
 RN (1)
 RP NUCLEOTIDE SEQUENCE.
 RC STRAIN=TT01;
 RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
 RA Duchaud E., Rusnlok C., Frangeul L., Buchrieser C., Gavaudan A.,
 RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
 RA Dassa E., Derose R., Derzelle S., Freyssinet G., Gaudriault S.,
 RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.,
 RT "The genome sequence of the entomopathogenic bacterium Photobacterium
 luminescens."
 RL Nat. Biotechnol. 21:1307-1313(2003).
 DR EMBL; BK571862; CAE13266.1; -; Genomic_DNA.
 DR PhotoList; plu0971; -
 KW Complete proteome.
 SQ SEQUENCE 2215 AA; 252782 MW; 188CEB6A61E69956 CRC64;
 Query Match 34.4%; Score 4491; DB 2; Length 2215;
 Best Local Similarity 39.8%; Pred. No. 8.2e-230;
 Matches 1012; Conservative 394; Mismatches 737; Indels 402; Gaps 47;
 QY 53 TIEQKNNRLLEARIFTTANPOLSGAIRLIGIERSV-SRSYDEMFGARSSSVFKPGSVAS 111

Qy	2216	YRRRQWEIQRDNADGEVKOMDAQLESLSKIRREAAQMVVEYQETQAQAELELQK	2275
Db	1891	YRRRQWEIQRNNAAEIKQIDAQLOSLAIRREAAVLQKTSLKQTQEQTAQLTFLOK	1950
Qy	2276	FTYKLSYWRGKLSAIYQFFDLTQSFCLMAQEAALRELTGNGVTIRGGANNGTAGL	2335
Db	1951	FSNQTLVYHMLGRUAAIFYDYDIAVARCLMAEYAEWETNTAANFIKPGAWQGTAGL	2010
Qy	2336	MAGETLLNLAEMEKVLERDERALEVTRTVSLAQFYQALSSDNFNLTQKLTQFLREGK	2395
Db	2011	LAGETLLMLNAQMEDAYLRWDQRALEVERTISLTQFYGALPEKSFNLATRISALLAGTT	2070
Qy	2396	NVGASGNELKLSNRQIEASTVRLSDLKTFSDYPE--SLGNTRQLKQSVTLPALVGPVEDI	2453
Db	2071	D-PVDHPPVTVLENDQLSAKLSLSLSLHSDYPEGNGVGNTRRIKQISVTLPALGPPQDV	2129
Qy	2454	RAVLNYGGSIV-MPRGCSAIALSHGVNDSQFMLDFNDSRYLPEGISVNDSSGLTUSFP	2512
Db	2130	QAILSYGGSSETGLAESCKLSAISGINDSQQLDFNNGKLLPEGIAINDTGILLNFP	2189
Qy	2513	DATDRQKALLESLSDIILHRYTIR 2537	
Db	2190	NATSKQAMQALSIIILHRYTIR 2214	
RESULT 13			
Q6XPA4 YERFR			
ID	Q6XPA4 YERFR	PRELIMINARY;	PRT; 2373 AA.
AC	Q6XPA4;		
DT	05-JUL-2004 (TrEMBLrel. 27, Created)		
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)		
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)		
DE	Putative insecticidal toxin TcYF1.		
GN	Name=tcYF1;		
OS	Yersinia frederiksenii.		
OG	Plasmid pMYE1.		
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;		
OC	Enterobacteriaceae; Yersinia.		
OX	NCBI_TaxID=29484;		
RN	[1]		
RP	NUCLEOTIDE SEQUENCE.		
RC	STRAIN=49;		
RA	Dodd S.J., Clare T.R., O'Callaghan M., Ronson C.W.;		
RT	"Yersinia enterocolitica contains plasmid-encoded genes that show		
RT	similarity to the tc family of insecticidal toxins.";		
RL	Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AY220302; AAP48861.1; -; Genomic_DNA.		
DR	InterPro; IPR006162; Poptant S.		
DR	PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN_1.		
KW	Plasmid.		
SQ	SEQUENCE 2373 AA; 263383 MW; BOBEECI0A1CC4568 CRC64;		
Query Match 31.1%; Score 4061; DB 2; Length 2373;			
Best Local Similarity 37.6%; Pred. No. 7.9e-207;			
Matches 999; Conservative 403; Mismatches 845; Indels 412; Gaps 55;			
Qy	1	MYSTAVLNKISPTRGQT-----MTLADLYLSFSELAKIFDDQLSGEAKHLVHETI	54
Db	6	MYRTDDILEKLNINERQSESVTAETLADFFSRSLSEVKKITGDNLWSGEVRYLYRQAQ	65
Qy	55	EQKNWELLERAFITRANPQLSGAIRLIGERDSVSRYSDEWFGARSSSVKPGQSVASMF	114
Db	66	KEQKENLTSRIILARANPQSGAARLIGIRQAADTRSYDEWFGSRAERFTRPQSVASMF	125
Qy	115	PAGYLITELYREAKDLHPSSSAYHLDNRRLADIATLSQSNMDTEISTLTLSNELLLEHIT	174
Db	126	PAAYQTELYREAKDLHPATSLYRLDNRDLAQALSQSNMDREITTLGLSSELLERSIG	185
Qy	175	RKTGGSDALMESLSTRQADITPYQETIROVIMTHDSTLSALSRNPEVMQAGAS	234
Db	186	AAGIGDVDSVRDLIAEYRLATGLSPDHTAY-----RRP-AKPFCAGAR	226

```
Db 1245 ---NTVLVSVDSDMKLNTVSDNLI--PQSIKHEFN-----TSTERFINNIFSD 1287
Qy 1317 ASLWNGSAIGDSDUTVMWGNIPQITSKYSS-----DNLAITLHNAAFVTRVGGSGNVIRN 1372
Db 1288 SSASYAIS-GTDSIDDATHSDILLNSRVMTVITDN-SMGSWNPEFNIAASVCFITNS 1345
Qy 1373 KQISAMKLTGVDGSKSYGNANFIANTVKHYGGYSDLAGGPITVYKNTKNIYASV---QGH 1428
Db 1346 SLVS-----DEYNYFI-----MGG---VFPERKNFYVAINWDRPGG 1379
Qy 1429 LWNADYTRRLILITPVNNYIARLFEFPSPNTILMTVTVGNSKTSDFK-KCSYAVDGN 1487
Db 1380 FIGVD-----VSN-----SKIYQVG--KASDFVIHKTSVSGVQ 1411
Qy 1488 SQ-----GQIFSSVQS-----SGWLDIDTGINNTDIKITYWAGSKTHFTASD- 1531
Db 1412 GQIELNVHSGKHLFGTLTDAMVAALLNGSTSTTSILN-QATITAGFAGVLPVNVANDL 1470
Qy 1532 -----HIASLPA-----NSPDAMPYTKPLEIDASSLAFTNNI 1564
Db 1471 IDDTLITARMNGTEIKSPSEWYNRELNVGNLNFNTALSF---TVDTSDI--TDDE 1525
Qy 1565 APLDIVFETKAGDGRVLGKIQTLTSVKRVNPNPDIILFRETSHSGAQMQL-----GYVRI 1620
Db 1526 FDVTLTFTAVGENDAVLAARTVINVRRLINNDTPVIALRKNTRGAQYIRFTAGNDVALI 1585
Qy 1621 RLNTLLASQLVSRANTGIDITLTMETORLPPPPLEGEPFANFVLPKYDPAEHDGDERWFKI 1680
Db 1586 RLNTLFARQLVDRANTGIDITLTSMETORLTPEALEEG----- 1622
Qy 1681 HIGNVGNGTGRQPYYSGLMSPDTSMTLFPYABGYMYHSGVLGVGYQKIYDNTWES 1740
Db 1623 -----SDVF----- 1626
Qy 1741 AFFYFDETKQOQVLINDADHDSGMTQOQIVKNIKKYKGLNVSATGYSPAPMDFNSASAL 1800
Db 1627 -----MDFSGANAL 1635
Qy 1801 YWELFYPTPMCFORLLOEQFBQATQWIVNVNPNAGYVINGBIAPWNCRPLEETTS 1860
Db 1636 YFWELFYPTPMVFORLLOEQFBQATQWIVNVNPNAGYVINGBIAPWNCRPLEETTS 1695
Qy 1861 WNANPLDAIDPDAVQNDPMHYKATFMRLLDQILRGDMAYRELTRDALNEAKWYVRT 1920
Db 1696 WNDSPLOSVDPAVAQNEPMHYKATFMRLLDQILRGDMAYRELTRDALNEAKWYVRT 1755
Qy 1921 LELLDGPEDEYGSQQAAPSLSGAASQTVQAAQQDLTMLGRGVGSKNLRNTANSVLGLFL 1980
Db 1756 LNLGGERPYISFDADWSALTIGDAASEVTARSYQSALLAVRRQQGAPDVRTANSITLTL 1815
Qy 1981 PEYNPALTDYQWTLRLFLPNLRHNLSDGQPLSLAIYABPTDPKALLTSWVQASQGGSAV 2040
Db 1816 PQQNAVLLKGYWQTLAQRLLYNLRLHNLSDGQPLSLAIYABPTDPKALLTSWVQASQGGSS 1875
Qy 2041 LPGTSLSYRFPVLMERTENVAQLTOFGTSLLSWAHDDADELTLLLQQGWEIATQSIR 2100
Db 1876 PVAWNPYRFPVLMENARGQVSLIQQFGSTLLSITERQDAELAEILLQTSQSEILLOGLR 1935
Qy 2101 IQORTVDEVDADIIVLAESRRSAQNRLKYOQLYDEIDINHGEOQRAMSILLDAAGQSLAGQ 2160
Db 1936 QQDNALAEIDADILALEESRMGAQARFEYYSRLYDADVNTKEQAMDLYLSSSVLSASSQ 1995
Qy 2161 VLSTABGVADLVNVFGLACGSGRWGAALRASVMSLSATASQYSDAKTISRSEAYRRRR 2220
Db 1996 VLFMAGAADMPLNPIYGLAVGSGRYGALFNATAIGIQVSSDATRISADKISQSEMYRRRR 2055
Qy 2221 QEWIQRDNAGEYKQMDAQLESKIRREAAQMOVEQETOQAQTOALELQKFTNKA 2280
Db 2056 EDWEIQRDGAQSDVTGMDAQLAVNVRREGAELOKTYNEMQOIQTOQAQMAVLOKFSNTA 2115
Qy 2281 LYSWMRGKLSAIYYQFFDLTOSFCILMAQEARLRELTDNGVTFFIRGGAWNGTTAGIMAGET 2340
```

```
Db 2116 LYSWLRGKLAALYYQFYDLTASRCLMAQSAQYWD-RNSTTRFIQPGAWQGTYAGLLAGET 2174
Qy 2341 LLLNLASWEKVMLEDERALERALEVTRTVSLAQFYQALSSD-NENLTEKLTQFLREGKGNVGA 2399
Db 2175 LMLGLSRMEQAWLESDBREREVTRTVCLSEVIAGLAGDAAFVLADDEVVGLVNGGTSSAGT 2234
Qy 2400 SGNELKLSNRQIEASVRLSDLKIFSDYPESIGNTRQLKQVSVTLPALVGPYEDIRAVLNY 2459
Db 2235 ATNGLKPADQOQLATLSADLNIRSDYPSLGGFTRIKQISVTLPALVGPYQDYRAMLSY 2294
Qy 2460 GGSIVMPRGSAIALSHGVNDSGQFMDFDNRVLPFEGISVNDGSLTTLSPDATDRQK 2519
Db 2295 GGSVMVPRGCNALAVSHGMNDSGQQLDFDNRVLPFEGIPVGDGSLTTLSPDADGRQ 2354
Qy 2520 ALLESLSDIILHRYTIRS 2538
Db 2355 AMLLSLSDIILHRYTIRS 2373

RESULT 14
Q4ZNN7_PSES
ID Q4ZNN7_PSESY PRELIMINARY; PRT; 2502 AA.
AC Q4ZNN7;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Insecticidal toxin complex protein Tcdal.
GN ORFNames=Psyr_4205;
OS Pseudomonas syringae pv. syringae B728a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205918;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RG DOE Joint Genome Institute;
RA Chain P., Larimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,
RA Nolan M., Goltsman E., Thiel J., Malfatti S., Lapidus A., Detter J.C.,
RA Land M., Richardson P.M., Kyrpides N.C., Ivanova N.;
RT "Comparison of two complete genome sequences of Pseudomonas syringae
RT pv. syringae B728a and pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Loper J.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Feil H., Feil W.S., Lindow S.E.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000075; AA39235.1; -; Genomic DNA.
SQ SEQUENCE 2502 AA; 277032 MW; A0069019F516BF0A CRC64;

Query Match 30.3%; Score 3948; DB 2; Length 2502;
Best Local Similarity 36.0%; Pred. No. 9.1e-201;
Matches 962; Conservative 474; Mismatches 922; Indels 314; Gaps 68;

Qy 8 LNKISPTRDGQMTLADLOVLSFSEL-----RKIFDDQLSWGEARHLYHETIEQKQN 60
Db 3 LNRLLS--RIKRLSARAGLSDLSAAQALISPAFRNKLSTSGLSPREVDLYESARDREA 60
Qy 61 RLLEARIFTRANPOLSGAIRLGIE-RDSVSRSYDEMFGARSSSVFKPGSVASMSPPAGYL 119
Db 61 LIYEKCLLARSSPLLNKAVRLGINPPDASLRDYEQFGRASSYTLPGSVSMFSPAAYL 120
Qy 120 TELYREAKDLHFSSAYHLNRRPDLDLTLTQSNMUTEISTLTLSNELLE----- 171
Db 121 SALYRNARGLYPESPPIHDKRRPDLKGLSQQNMKEVSALSLSNEVMTLAGESLKL 180
Qy 172 HITRKTGSDSALMESLSTYRQAIDTTPHOPYETIRQIVIMTHDSTLSALSNNPEVMGOAE 231
```


Db 181 DVTAEGQG---SVLEWLSYTLRSGSTPYHHPHARLRQSOIQKDPKPKOLAANPRVTGLFS 237
Qy 232 GASLAILANISPELYNLTBEITEKNADALFAQNFSENIITPENFASQSWIAKYGLBLS 291
Db 238 GATWAGMADISPEYALITEEVTAGSADLEYARNFG-TIDPETLITPQSLRRYGLSDE 296
Qy 292 EVQKYLQMLQNGYSYSTSAYVDNIISTGLVNNESKLEAYKTRVKTDYDDKNIYFDLMY 351
Db 297 EVALFSTDYANG-BEGABEYINNLVTRIGD-----KIYRQSGHPTNTLAWLLP 347
Qy 352 EGNQOPFRANPKVSREFGATLRKNAGSPGIVGSL-----SGPLANTNPK-----SNY 400
Db 348 QADGTWLLKSF-----SDAQGYDIFQVRVDTADAIYEENQEMGDFVPGQSY 396
Qy 401 LSNISDSSEYKNG---VKIYAYRYSTSTSATNQGGGIFTFESYPLTIFALKLNKAIRLCIT 457
Db 397 TVSIEASDLLQGEFTMSLRHYNGSYVSGVSYNNVIIDSF--QFYVLKLNKAIRLWRA 454
Qy 458 SGLSPNELQTVRSNAGIINDSVLTKVFTLFLYSHRYALSPDDAQVLNGSVINQYADD 517
Db 455 TGLHPKSLVIVNSVPVN-ITDETQLLLFQVQRCVQRYGVEPEALVLCGLLSNNSYD 513
Qy 518 DSVSHFRLNTPPLPKGKIFEADGNTVSIIDP-DEEQSTFARSALMRGLGVNSGELYQLGK 576
Db 514 DNQSLFDQVNSPPLNGETIAPTSTISINLLPTNATDNTVEKAILKRALNDDVGLFALLR 573
Qy 577 LAGVLDQNTITLSVFVTSLSYRLTLARVHQLTVNELCMLYGLSPFNKTKTASISGEL 636
Db 574 IFDNAASSGVLTNLKNLSAMVALSRWALHALSVTELEQL-----LKAAGL 620
Qy 637 PRLVI-----WLYQV---TOWLFEAITTEAIWLLCTPBFSGNISPEISNLLNLL 683
Db 621 PRLTSELESTQLWSGLQKVDLSLTQWLSNRKLSVADLELLTRPTPTQAAATEISALLEEL 680
Qy 684 RPSISEDMAQSHRELQAEIAPFAATLHLASPDARYIILLTNDTPGGLDIAGMTL 743
Db 681 KGVISAH-PDADTLEKRLTALLPVLVSSALPSPAPVAESVLAWANGLOPABEWVDQFW-- 737
Qy 744 VLKESLNANET--TOLVOFCHVMAQLSLSVQTLRSEALSVLVTSIGFAVLGAKNQPAQO 801
Db 738 -----DGAETDDSDSVAPCYGLAQLALIYHATGINPOAFSLFVATPARLLG----PAPE 787
Qy 802 ---HNITDLPFLSRPHQWINGNPGSDTDLMLRQOTLTPADRLASVWGLDIDSMVTQAM 856
Db 788 VAVLPRSLATVQALCNFSAWLSKLDGASVLAAPFVADTLTPADLARALNDDAFREQAT 847
Qy 857 VSAGVN-----QLOCHQDINTVLQWIDVASALHTMFSVIRTVNIRYVTAALKAESNL 909
Db 848 ROAVANTOAASETKLASWSEIDVVLQW-----AALSTTFGV--TPVNIQGLLALNTAGNQ 901
Qy 910 PSWDEWOTLAENMEAGLSQOQAOTLADYTAERLSVLCNWFELANIQPBGVSLHSRDDLVS 969
Db 902 PPWDDWVRVADAFAGLTPTETKMEALASGLSAAALCYLLKSGMTAQLANNSREGLIQ 961
Qy 970 YFLIDNQVSSAIKTRRLAEIAGIQLYINRALNRIEPNARADVS--TRQFTDW-TVNRR 1026
Db 962 YLLLDNLNGPQWTSRVAEIVSVQTFIORTLSAAESQGTVDKAAVTGQFFTDWERYNQ 1021
Qy 1027 YSTWGGVSRUYVYPENYIDPTQRTGOTRMDDELLENISQSLSRDVEDAFKTYLTRPET 1086
Db 1022 YSTWAGAAKVVYPENYIDPTVRLGQSGMMNTMLQTLGQALNTDVTGDAFNTYLSNSEE 1081
Qy 1087 VADLKVSAYHDNVNSNTGLTFWGVOTRENLPYYWRNVDS-RMOAGELAAANAKWTK 1145
Db 1082 VANLRVSGYHDNLDVHEGKTYFFIGTNSQSEVREFFYRSADGRSEDQGLAANATDWRK 1141
Qy 1146 IDTAVNPKDAIRPVIFRERLHLIHWKEBEVA---KNGT--DPVETYDRFTLKLAFLRHD 1200
Db 1142 IECAAPWGDCTIRPVYKSRSLYVLCWLERKDVTPPKADGSPGD-KMFD-YAINISYLRD 1199
Qy 1201 GSWAPSWYDITTOVE--AVTDKXPDTERLALASFGQEDTLLVVFYKTKGYSDFGGS 1258
Db 1200 GNWTSPIIIDVTEENGPAAGKPP-----GIYCTSFDAETTMVALLYRKADSVADQLPV 1254

Qy 1259 NKXVAGMTIYDGSFKKMENTALSRYSQLKNTFDIIHTQOGLNDLVRKASRYFAQDFEVPAS 1318
Db 1255 QADMELVVIYEDMASDQSSAETFLSSMK--YELDPTTEGNVWNNKYTH----- 1301
Qy 1319 LMGSAIGDDSLTWENGINIPOITSKYSSDNLAITLHNAAFTRYVDSGNVIR---NKOI 1375
Db 1302 ---GLSLGSESVHGLYERNY-DISVSYS-----YTLVNGSEAVAKISIRKI 1346
Qy 1376 SAM-----KLTGVODG-----SOVGNAFIIANTVKHYGGYS 1407
Db 1347 TLVPRIPNPLRLLYEYSSPAETFFLRLASDDGKIWFVROGGKIYLLSLTILTW----- 1401
Qy 1408 LGGPITVYNTKKNYIASVOGLHMANADYTRRLILITPEVNNYIARLPEFPFSPNTILNTVPT 1467
Db 1402 -GVPLSVM-----VYGDHARLLVR--EDAVSAVTF-----VLA 1433
Qy 1468 VGSNKTSPDKCSYAVDGNNSQGFIF-----SSVQ-----SSGWL-----DIDTGIN 1510
Db 1434 VYNDPFDMLCP-----SPAGFTVMIQLVTGYSQFDTPPQPMANPPYFSDL----- 1482
Qy 1511 NTDIKITVMAGSKTHTFTASDHIASLPANSPDAMPYTFKPLEIDAS-----SLAFT 1561
Db 1483 ---LQNLVFDVNGQAFLAGDYVSALPPFDISGMVDFDVSNSISKEWGLVPSYDVAFS 1539
Qy 1562 NNIAPLDIVFETKAKGRVLGKIKOTLSVKRVNYPEDILFLRTHSGAQYMLGVYRIR 1621
Db 1540 PSIGVD-----SDDR---IQWIVPVRNASDSKNIIKIKNTPEKAYMEPDACRTR 1588
Qy 1622 LNTLLASQLVSRANTGIDTILTMETORLPBPPLGEGFPANFVLPKYDPAEHGDEMFH 1681
Db 1589 LNTLFARQLVERAAGIDTILSYETOIEQEPOLGAGFTTLTLPFHKGVHGDEPMVKIY 1648
Qy 1682 IGNVGNTGRQPYYSGLMSDTSMTLFPVYAE-GYVMEHGVRLGVGYKITDY--NTW 1738
Db 1649 XNYFLAVNDSYLANSGSLSATSTTVNLFIYPDGGFTTVEVHLRLQYKEADYNGLSQ 1708
Qy 1739 ESAPFYFDETKQOPVLINDADHDSGMTQOQGIIVKNIKKYKGFNLVSIATGYSAPOFNSAS 1798
Db 1709 SWLLYAPDTGVATV---SRPGTNSLSEIVK-----SVTVKSRSSTVPMDFSGAN 1755
Qy 1799 ALYVWELFYTPMCMFORLLOEKOPDEATQINVYVNPAGYIVNGEIAPIWNCPLPET 1858
Db 1756 ALYVWELFYTPMMAAORFLOEQOFTLADPRLYVSPSGYVVRGQHVDRNWNVPLOED 1815
Qy 1859 TSWANPLDADPDPAQNDPMHYKIATFMELLQDLILRGDMAYRELTRDALNEAKMYV 1918
Db 1816 TSWNDAPLKAVDPDAQNDPMHYKVATFMERALDILLIARGDAAYKLERDYLAEAKVYS 1875
Qy 1919 RTLELLGDEPEDYSGQWAAAPSLSGAASQTVQAAAYQOBLTWM-----GRGGVSVKNLRTAN 1973
Db 1876 QALNLLGEPQYIRANALWAEPSLGEASS---EALAEQHLTVLSLLRDRGVETLKSMASTK 1932
Qy 1974 SLV--GLFLPEYNPALTDYQOTLRLFLNLRHNLSDGOPSILAIYAEPTOPKALLTSMV 2031
Db 1933 TAAASALFLPEINVMQGYMLTLRQMYNLRHNLTDGQOPLLLFLFAKPAKALLNAV 1992
Qy 2032 QASQGSVAIPGT--LSLYRPPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTTLLQQ 2090
Db 1993 ASESSESSDLPVTSPLWRFEPLDSDAGLVFQLIQFNAVGLVLERQDAESLALLQ 2052
Qy 2091 GMEIATOSIRIQORTVDEVDADIADVLAESRSQAQRLEKYQOQLYDEDEINHGQRAMSLD 2150
Db 2053 GTELMASTIQOEGTRELBAEKVLSRSKDSQRRPDSYRSLYDEDEINRERLSSISLE 2112
Qy 2151 AAQOSLAGQVLSIAGVADLVPNVFGCACGSRWGAALRASASVMSLSATASQVSADKI 2210
Db 2113 SAKSLGAKAAYIAAALDLPAPNI FGLANGMKGFGVGTWAGLGISISAENLMDSSRI 2172
Qy 2211 SRSEAYRRRREWEIQRDNADGEVKQMDAQLESKIRREAAQOMOVEYQETQQAHTQAOLE 2270
Db 2173 SQEIIYRRRREWEIQRNAGEIQQIETAOQLASLEVRRESTELQXAHLEMOQQAQO 2232

Db 1159 YEEYONY--EDNRVLYNGSRDVKVYISAPVPMKSNFDEBERFNKLEKRNKYSLSFA 1216
QY 1454 PPFSENTILNTVFTVGS-----NKTSDFKKCSYAVDGN-----NSQGFQ 1492
Db 1217 GDYGNRHSHTLPSIGSKPNELIFAYVNDLLDEDLIEAYIDDNLFMGWSMWRNSDSYQ 1276
QY 1493 IF-----SSVQSSGWLDDIDGINNTDIKITVMAGSKTHPTASD-HIASLPANSP 1541
Db 1277 YYPEKNITITPTKTPASIESLITTSVGNKVV-----SGPYPTANNKKSADILVHKW 1330
QY 1542 DAMPYTPKPLEIDASSLAFTNNIAPLDIVPETKAKDGRVLGKI KOTLSVKRVNYPEDIL 1601
Db 1331 DOPIELTLPVASEST-----DIEIHSYHSRNEVSGSMIOTIPIITVDPVSQNYL 1382
QY 1602 FLRETHSGAQYMLG---VYIRIRLNTLLASQLVSRANTGIDITLMTQRLPEPPLGEGF 1658
Db 1383 ALKTNEAQAYLEWDKQKHQVRLNTLFAKELVREADSGIDNILSPDTQLIFEPDLGEN- 1441
QY 1659 FANFVLPKYDPAEHGDERWFKIHI GNVGNGTGRQPYSGMLSDTSETMTLFPVYAEYY 1718
Db 1442 ----- 1441
QY 1719 MHEGVLGVGYKITDYDNTWESAPFYFDETKQFVLINDADHDSGMTQGGIVKNIKYYKG 1778
Db 1442 -----NNHAAQOS----- 1449
QY 1779 FLNVSATGYSAPMDFNSASALYIWELFYTPMCMFORLQEKQFDEATOMINYVNPAG 1838
Db 1450 -----VDFSGANGLYFWELFYTPMLVMQRLQEQSPDLATHMLSYVFN- 1494
QY 1839 YVNGEIAPIWNCPLBETTSWNPANPLDADPDVAQNDPMHYKIATFMRLLDQLILRG 1898
Db 1495 ---TNPSTP--WNTRPLKEDTAWNAPLDSTNPDAVAQADPMHYKLSFTFMBLLDILLIARG 1549
QY 1899 DMAYBELTRDALNEAKMAYVTRLELLGDEP-EDYGSQQWAAAPSLSGAASQTVOAAYQDL 1957
Db 1550 DYAYHQOQRDITLAEAMMYVQNNLLGGEPTRTDMSRW----- 1588
QY 1958 TMLGRGGVSKNLRNTANSLVG-LFLPEYNPALTDYQTLRLRLNLRNLSIDGQPLSLAI 2016
Db 1589 -----VHLPLKEAVELADHLFLPQONTKLQDYQILKLRVFNLRNLSIDGQPLMLPQ 1641
QY 2017 YABPTDPKALITSMVOASQGSVALPG--TLSLYFPVWLERTRNLVAQLTQFGTSLSM 2074
Db 1642 FTAPADAKALMNTAVASSLGGTA-LPSHIKLSLQRFPPVLESAMVGVQLIQFGSSLSGV 1700
QY 2075 AEHDDADELTLLLOQGMELATQSTRIQORTVDEVDADI AVLAEBSRRAQNRLEKYQOLY 2134
Db 1701 IERDAEKMAVLMQGTGTELLAQSVLMQOKNLEELHETHTALSALAGAEGRYKHNDLY 1760
QY 2135 DEDINHGQRAMSLDDAAGOSLACQVLSIAEGVADLVPNVFGACGSGRWGAALRASAS 2194
Db 1761 QENINAGENSALRATAGFITGTTQALYMTGAALDMVPNIYGVMSVGARYGATANAJAI 1820
QY 2195 VMSLSATASQYSADKISRSEAYRRRRQBEWEQORDNADGEVQMDAQLESKIRREAAQMQ 2254
Db 1821 GSSIASTASTLAADGISTTEMYRRRRQBEWEISRNNABAEIKQITAAQETLKIRSTSAQLQ 1880
QY 2255 VEYQETQQAHTQAQLELLQKFTNKALYSWMRGKLSALYYOFFDLTQSFCLMAQEARRE 2314
Db 1881 VSYLETQQAQHTQALMNTKFTNQALYSWMRSRLSALYFQYDLTUSRCLOAQYGYOWE 1940
QY 2315 LTDNGVTPIRGAMNGTTAGLMAGETLLINLAEMEKVWLERDERALEVTRTVSLAQFYQA 2374
Db 1941 TLDI-TSFIKPGAWQTYAGLLCGEALMLNLTLEAAVQSMQARELEIERTLSLAELYQN 1999
QY 2375 LSSDNFNUTEKLTQPLREGK--GNVGSAGNELKLSNRQIBASVRLSDLKIFSDYPE--SL 2430
Db 2000 IPTSGFDLATAISAALSASNKVSQTCENGNVISIKDDILSLSFISALNLAQDYPKIMGL 2059
QY 2431 GNTROLKQVSVTLPALVGPYEDIRAVANYGGS-IWMPGCSAIALSHGVNDSGGOFMLDPN 2489
Db 2060 GDKRPFQIQISVSLPALLGYSQVAVLSYTGTDATFAFGCOAIALSGMNDSGLPQLDPN 2119

QY 2490 DSRYPPEGISVNDGSLTSLSPDATDRQKALLESLSDIILHIRYTIRS 2538
Db 2120 DSNYLPPEGIDIEDSGFVLRFPTNSDKQKALQSLSDIILHIRYTIRS 2168

Search completed: February 16, 2006, 21:42:47
Job time : 366.476 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:22:44 ; Search time 251.656 Seconds
(without alignments)
4431.220 Million cell updates/sec

Title: US-10-754-115-34
Perfect score: 13043
Sequence: 1 MYSTAVLKNKISPRDGTQM.....KALLLESLSIIILHRYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21:.*
1: Geneseqp1980s:.*
2: Geneseqp1990s:.*
3: Geneseqp2000s:.*
4: Geneseqp2001s:.*
5: Geneseqp2002s:.*
6: Geneseqp2003as:.*
7: Geneseqp2003bs:.*
8: Geneseqp2004s:.*
9: Geneseqp2005s:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	13043	100.0	2538	8	Adn61385 Xenorhabd
2	13043	100.0	2538	8	Adr20371 Recombina
3	13043	100.0	2538	8	Adr21514 Xenorhabd
4	13043	100.0	2538	8	Adr21500 Xenorhabd
5	5496.5	42.1	2523	8	Adr20365 Recombina
6	5496.5	42.1	2523	8	Adr21494 Xenorhabd
7	5148.5	39.5	2522	2	Aay33729 Photorhab
8	5043	38.7	2516	2	Aaw56572 Toxin Tcd
9	5043	38.7	2516	4	Aab72609 Photorhab
10	5043	38.7	2516	5	Abg32651 P. lumine
11	5043	38.7	2516	8	Adp18614 Photorhab
12	5043	38.7	2516	8	Adr21575 Photorhab
13	5043	38.7	2516	9	Adr21517 P. lumine
14	5043	38.7	2517	4	Aab72611 Modified
15	5043	38.7	2537	4	Aab72614 TcdA toxi
16	5003	38.4	2516	2	Aaw17899 Photorhab
17	4986.5	38.2	2504	2	Aaw56557 Toxin tcb
18	4986.5	38.2	2504	4	Aab72610 Photorhab
19	4986.5	38.2	2504	5	Abg32654 P. lumine
20	4986.5	38.2	2504	4	Adr21539 Photorhab
21	4986.5	38.2	2505	4	Aab72612 Modified
22	4976	38.2	2534	9	Aeb47821 Native Xp
23	4966.5	38.1	2504	2	Aaw17871 Photorhab
24	4914.5	37.7	2526	6	Abm70229 Photorhab

25	4802	36.8	2381	8	ADP18620	Adp18620 Photorhab
26	4802	36.8	2381	8	ADR21543	Adr21543 Photorhab
27	4763.5	36.5	2466	6	ABM70225	Abm70225 Photorhab
28	4583.5	35.1	2376	4	AAy97694	AAy97694 SepA prot
29	4574	35.1	2499	8	ADP18616	Adp18616 Photorhab
30	4572	35.1	2499	8	ADR21542	Adr21542 Photorhab
31	4520.5	34.7	2499	6	ABM70221	Abm70221 Photorhab
32	4491	34.4	2222	6	ABM70220	Abm70220 Photorhab
33	4433	34.0	2177	8	ADP18618	Adp18618 Photorhab
34	3303	25.3	1849	2	AAW56573	AAW56573 Toxin Tcd
35	3278.5	25.1	1844	2	AAW18302	AAW18302 Photorhab
36	3278.5	25.1	1844	2	AAW56558	AAW56558 Toxin Tcb
37	3264	25.0	1849	2	AAW17900	AAW17900 Photorhab
38	2855	20.4	1273	9	ABE47816	ABE47816 Partial n
39	1643	12.6	1205	8	ADN61341	Adn61341 Paenibaci
40	1632	12.5	573	2	AAW56559	AAW56559 Toxin Tcb
41	1622	12.4	573	2	AAW18303	AAW18303 Photorhab
42	1615	12.4	579	2	AAW56574	AAW56574 Toxin Tcd
43	1611	12.4	579	2	AAW18301	AAW18301 Photorhab
44	1605.5	12.3	1370	6	ABM67445	Abm67445 Photorhab
45	1595.5	12.2	1565	6	ABM69065	Abm69065 Photorhab

ALIGNMENTS

RESULT 1

ADN61385
ID ADN61385 standard; protein; 2538 AA.

XX AC ADN61385;

XX DT 01-JUL-2004 (first entry)

XX DE Xenorhabdus nematophila Xwi XptA2 toxin complex protein.

XX KW Cry; toxic; lepidopteran pest; toxin complex; insecticide; xwi; xptA2.

XX OS Xenorhabdus nematophila.

XX PN WO2004002223-A2.

XX PD 08-JAN-2004.

XX PF 27-JUN-2003; 2003WO-US020082.

XX PR 28-JUN-2002; 2002US-0392633P.

XX PR 21-JAN-2003; 2003US-0441647P.

XX PA (DOWC) DOW AGROSCIENCES LLC.

XX PI Bintrim SB, Bevan SA, Zhu B, Merlo DJ;

XX DR WPI; 2004-082821/08.

XX PT Screening a culture of Paenibacillus isolate for Cry protein or toxin complex protein, useful for controlling lepidoptera, comprises obtaining DNA or protein from the culture and assaying the presence of the gene or protein.

XX PS Example 10; SEQ ID NO 49; 220pp; English.

XX CC The invention relates to a novel method for screening a culture of a Paenibacillus isolate for a gene encoding a protein selected from a Cry protein that is toxic to a lepidopteran pest and a toxin complex protein. The method comprises obtaining DNA from the culture and assaying the DNA for the presence of the gene or obtaining a protein produced by the culture and assaying the presence of a protein that indicates the presence of the gene in the isolate. The method of the invention has insecticide applications and may be useful for screening Paenibacillus sp. for toxin complex (TC)-like genes and proteins which may themselves be used to enhance or potentiate the activity of a stand-alone Xenorhabdus toxin protein. The method may also be useful for screening

CC Paenibacillus sp. and others for insecticidal thiaminase genes and
CC proteins for controlling insects, particularly lepidopterans. The current
CC sequence is that of the xenorhabdus nematophila xwi XtpA2 toxin complex
CC protein of the invention.

1	QY	1	MYSTAVLLNKISPTPDGQTMTLADLQYLSFSSELKIPDDQLSWGSGARHLYHETIIBQKKN	60
2	DB	1	MYSTAVLLNKISPTPDGQTMTLADLQYLSFSSELKIPDDQLSWGSGARHLYHETIIBQKKN	60
3	QY	61	RLLEARIPTRANPOLSGAIRLGIERSDSVRSYDENFMFGARSSSFVKPFGSVASMPSPAGYLT	120
4	DB	61	RLLEARIPTRANPOLSGAIRLGIERSDSVRSYDENFMFGARSSSFVKPFGSVASMPSPAGYLT	120
5	QY	121	ELYREAKDLHFSSSAYHLDNRPPDLADLTLSQSNMDTEISTLTLSNELLLEHITRTKGGD	180
6	DB	121	ELYREAKDLHFSSSAYHLDNRPPDLADLTLSQSNMDTEISTLTLSNELLLEHITRTKGGD	180
7	QY	181	SDALMESLSTYRQALDTPYHQPYETIROVIMTHDSTLSALSRNPVWMOAEGASLLAILA	240
8	DB	181	SDALMESLSTYRQALDTPYHQPYETIROVIMTHDSTLSALSRNPVWMOAEGASLLAILA	240
9	QY	241	NISPELYNILTEETEKNA DALFAONFSENITPENFASQSWITAKYGLSELSEVQKYLGM	300
10	DB	241	NISPELYNILTEETEKNA DALFAONFSENITPENFASQSWITAKYGLSELSEVQKYLGM	300
11	QY	301	QNGYSDSTSAVDNISTGLVWNESKLRAYKTRVKTDDYDKNINFDLWYEGNNQFPFR	360
12	DB	301	QNGYSDSTSAVDNISTGLVWNESKLRAYKTRVKTDDYDKNINFDLWYEGNNQFPFR	360
13	QY	361	ANFKVSRFEGATLRKNAGPSGIVGSLGPIANTNFKSNLYLSNISDSSEYKNGVKIYAVRY	420
14	DB	361	ANFKVSRFEGATLRKNAGPSGIVGSLGPIANTNFKSNLYLSNISDSSEYKNGVKIYAVRY	420
15	QY	421	TSSSTSATNOGGGIFTFESYPLTIFALKNKARLCITSGLSPNELOTVRSNDAQIIND	480
16	DB	421	TSSSTSATNOGGGIFTFESYPLTIFALKNKARLCITSGLSPNELOTVRSNDAQIIND	480
17	QY	481	SVLTKVFFYTLFYSHRYALSFDQAQVLNGSVINQYADDDSVSHFNLFNTPPLKGKIFEAD	540
18	DB	481	SVLTKVFFYTLFYSHRYALSFDQAQVLNGSVINQYADDDSVSHFNLFNTPPLKGKIFEAD	540
19	QY	541	GNTVSIIDPDEQSQTFARSALMRGLGVNGSELYQLGKLAGVLDAQNTILTSVFIISLRYL	600
20	DB	541	GNTVSIIDPDEQSQTFARSALMRGLGVNGSELYQLGKLAGVLDAQNTILTSVFIISLRYL	600
21	QY	601	TLLARVHQLTVNELCMLYGLSPFNKTTASLSGSELPRLVIWLQVQWMLTEAEITTEAI	660
22	DB	601	TLLARVHQLTVNELCMLYGLSPFNKTTASLSGSELPRLVIWLQVQWMLTEAEITTEAI	660
23	QY	661	WLLCTPEPSGINSPEISNLLNLRPSISDQMAQSHNRELQAEILAPPTAATHLASPDMA	720
24	DB	661	WLLCTPEPSGINSPEISNLLNLRPSISDQMAQSHNRELQAEILAPPTAATHLASPDMA	720
25	QY	721	RYIILLTNDLNRPGGLDIAGFWTLVLKESINANETTLQVQFCHVMAQLSUSVQTLRLSEAE	780
26	DB	721	RYIILLTNDLNRPGGLDIAGFWTLVLKESINANETTLQVQFCHVMAQLSUSVQTLRLSEAE	780
27	QY	781	LSVLVVISGFAVLGAKNQAGQHNIDTLFSLYRPHQWINGLGNPGSDTLDMLRQOQTLTADR	840
28	DB	781	LSVLVVISGFAVLGAKNQAGQHNIDTLFSLYRPHQWINGLGNPGSDTLDMLRQOQTLTADR	840
29	QY	841	LASVWGLDISHVWTQAMVSAGVNOLOCQDINTVLQWIDVASALHTWPSVIRTLVNIIRYVT	900
30	DB	841	LASVWGLDISHVWTQAMVSAGVNOLOCQDINTVLQWIDVASALHTWPSVIRTLVNIIRYVT	900
31	QY	901	ALNKAESNLPWSDEWQOTLAENNEAGLSGTOAQTLADYTAERLSVLCNWFANIIOPEGVS	960

QY 2041 LFGTSLYRFPVLMRTRNLVAQLTQFGTSLLSMAEHDDADDELTTLLQQGMELATQSIR 2100
 DB 2041 LFGTSLYRFPVLMRTRNLVAQLTQFGTSLLSMAEHDDADDELTTLLQQGMELATQSIR 2100
 QY 2101 IQQRTVDEVDADIAVLASRRSAQRLEKYQQLYDEDINHGQRAMSLDDAAAGOSLAGQ 2160
 DB 2101 IQQRTVDEVDADIAVLASRRSAQRLEKYQQLYDEDINHGQRAMSLDDAAAGOSLAGQ 2160
 QY 2161 VLSIAEGVADLVPNVFGACGSRGGAALRASASVMSLSATASQYSADKISRSEAYRRR 2220
 DB 2161 VLSIAEGVADLVPNVFGACGSRGGAALRASASVMSLSATASQYSADKISRSEAYRRR 2220
 QY 2221 QEWEIFQDNADGEVQMDAQLESKIRREAAQMQVEYQETOQAHTQAQLELLQRKFTNKA 2280
 DB 2221 QEWEIFQDNADGEVQMDAQLESKIRREAAQMQVEYQETOQAHTQAQLELLQRKFTNKA 2280
 QY 2281 LYSWNRGKLSAIYYQFDLTQSFCLMAQALRELTNDGVTPIRGAWNGTTAGLMAGET 2340
 DB 2281 LYSWNRGKLSAIYYQFDLTQSFCLMAQALRELTNDGVTPIRGAWNGTTAGLMAGET 2340
 QY 2341 LLLNLAEKWKWLERDERALEVTRTVSLAQFYQALSSDNFNLTETKLTQFLREGKGNVGAS 2400
 DB 2341 LLLNLAEKWKWLERDERALEVTRTVSLAQFYQALSSDNFNLTETKLTQFLREGKGNVGAS 2400
 QY 2401 GNEKLKSNRQIEASVRLSDLKIFSDYPSLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG 2460
 DB 2401 GNEKLKSNRQIEASVRLSDLKIFSDYPSLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG 2460
 QY 2461 GSIWVPRGCSAIALSHGVNDGQFMDLDFNDSRYLPFEGISVNDGSLTSLSPDATDROKA 2520
 DB 2461 GSIWVPRGCSAIALSHGVNDGQFMDLDFNDSRYLPFEGISVNDGSLTSLSPDATDROKA 2520
 QY 2521 LLESLSDIILHIRYTIIRS 2538
 DB 2521 LLESLSDIILHIRYTIIRS 2538

RESULT 2
 ADR20371
 ID ADR20371 standard; protein; 2538 AA.
 XX AC ADR20371;
 XX DT 04-NOV-2004 (first entry)
 XX DE Recombinant pDAB2097 cosmid open reading frame 7 protein SeqID 20.
 XX KW toxin complex; TC; insect; pest control; exochitinase; insecticide;
 XX KW pesticide; agriculture.
 XX OS Xenorhabdus nematophila.
 XX PN WO2004067750-A2.
 XX PD 12-AUG-2004.
 XX PF 07-JAN-2004; 2004WO-US000314.
 XX PR 21-JAN-2003; 2003US-0441717P.
 XX PA (DOWC) DOW AGROSCIENCES LLC.
 XX PI Bintrim SB, Mitchell JC, Larrinua IM, Apel-Birkhold PC, Green SB;
 XX PI Schafer BW, Bevan SA, Young SA, Guo L;
 XX DR WPI; 2004-581006/56.
 XX DR N-PSDB; ADR20370.
 XX PT New Xenorhabdus toxin complex proteins and polynucleotides, useful for
 XX PT controlling pests and insects or for intoxicating or poisoning infected
 XX PT insects.

PS Claim 1; SEQ ID NO 20; 153pp; English..
 XX This invention relates to novel nucleic acid molecules isolated from
 CC Xenorhabdus strain xwi that encode toxin complex (TC) proteins.
 CC Specifically, it refers to proteins that are toxic to insects and other
 CC pests, such that they can be used to control pests in an agricultural
 CC environment to reduce crop loss and improve crop quality. The present
 CC invention further describes an exochitinase enzyme obtainable from the
 CC xwi strain that can be used to intoxicate or poison infected insect
 CC populations. As such, these toxins function as orally active insect
 CC control agents that can disrupt or deter insect growth and/ or feeding,
 CC which may or may not cause death of the insect but will work to avoid
 CC insect-induced plant damage. The toxin is delivered via transgenic plant
 CC expression, formulated protein compositions, sprayable protein
 CC compositions or a bait matrix. Accordingly, these compositions exhibit
 CC insecticide and pesticide activities. This polypeptide sequence is a
 CC Xenorhabdus nematophila open reading frame protein sequence encoded by
 CC DNA from the recombinant pDAB2097 cosmid insert given in an
 CC exemplification of the invention.
 XX Sequence 2538 AA;
 SQ
 Query Match 100.0%; Score 13043; DB 8; Length 2538;
 Desc Local Similarity 100.0%; Pred. No. 0;
 Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYSTAVLANKISPTRDGQTMTLADIQYLSFSELRKIFDDQLDSWGEARHLYHETIEQKKN 60
 DB 1 MYSTAVLANKISPTRDGQTMTLADIQYLSFSELRKIFDDQLDSWGEARHLYHETIEQKKN 60
 QY 61 RLLEARIITRANPQLSGAIRLIERDSVRSYDEMFGARSSSFVKPGSVASFPAGYLT 120
 DB 61 RLLEARIITRANPQLSGAIRLIERDSVRSYDEMFGARSSSFVKPGSVASFPAGYLT 120
 QY 121 ELYREAKDLHFSSSAYHLNRRPDLADLTLSQSNMDTETISTLTGNEELLEHTTKTGSD 180
 DB 121 ELYREAKDLHFSSSAYHLNRRPDLADLTLSQSNMDTETISTLTGNEELLEHTTKTGSD 180
 QY 181 SDALMESLSTYRQAIDTPYHQPETIRQVIMTHDSTLSALSNNPEVMQAGASLLAILA 240
 DB 181 SDALMESLSTYRQAIDTPYHQPETIRQVIMTHDSTLSALSNNPEVMQAGASLLAILA 240
 QY 241 NISPELYNLTETETEKADALFAQNFSENITPENFASQSWITAKYGLSEVQKYLGM 300
 DB 241 NISPELYNLTETETEKADALFAQNFSENITPENFASQSWITAKYGLSEVQKYLGM 300
 QY 301 QNGYSDSTSAYVDNISTGLVNNESKLEAYKITRVKTDYDKNINVFDMYEGNQOFFIR 360
 DB 301 QNGYSDSTSAYVDNISTGLVNNESKLEAYKITRVKTDYDKNINVFDMYEGNQOFFIR 360
 QY 361 ANFKVSRFEGATLRKNAGPSGIVGSLGPLIANTNFKSNYLSNISDSSEYKNGVKIYARY 420
 DB 361 ANFKVSRFEGATLRKNAGPSGIVGSLGPLIANTNFKSNYLSNISDSSEYKNGVKIYARY 420
 QY 421 TSSTSATNQGGGIPTFESYPLTIFALKNKAIRCLTSGLSNPELOTVRSNAQGIIND 480
 DB 421 TSSTSATNQGGGIPTFESYPLTIFALKNKAIRCLTSGLSNPELOTVRSNAQGIIND 480
 QY 481 SVLTQVFTLTFYSHRYALSFDQAQVINGSVINQYADDDSVSHFNRLNTPPLKGIFFAD 540
 DB 481 SVLTQVFTLTFYSHRYALSFDQAQVINGSVINQYADDDSVSHFNRLNTPPLKGIFFAD 540
 QY 541 GNTVSDIDPEQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL 600
 DB 541 GNTVSDIDPEQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL 600
 QY 601 TLLARVHQLTVNELCMLYGLSPFNKGTASLSSGELPLRLVILYQVQWLTAEITTEAI 660
 DB 601 TLLARVHQLTVNELCMLYGLSPFNKGTASLSSGELPLRLVILYQVQWLTAEITTEAI 660
 QY 661 WLLCTPEFSGNISPEISNLLNLRPSISDWAQSHNRELQAEILAPFTAATLHLSPDMA 720
 DB 661 WLLCTPEFSGNISPEISNLLNLRPSISDWAQSHNRELQAEILAPFTAATLHLSPDMA 720

Qy	721	RYILLWTNLRPGGLDIAGFMTLVLKESINANETTQLVQFCHVMAQLSLSVQTLRLSEAE	780
Db	721	RYILLWTNLRPGGLDIAGFMTLVLKESINANETTQLVQFCHVMAQLSLSVQTLRLSEAE	780
Qy	781	LSVLVISGFAVLGAKNQAGQHNIDTFLSLYRFHQWINGLNGPGSDTLDMRLQOQTLTADR	840
Db	781	LSVLVISGFAVLGAKNQAGQHNIDTFLSLYRFHQWINGLNGPGSDTLDMRLQOQTLTADR	840
Qy	841	LASVGLDLSMVTQAMVSAGVNLQCCQDINTVLQWIDVASALHTMPSVIRTLVNIYYT	900
Db	841	LASVGLDLSMVTQAMVSAGVNLQCCQDINTVLQWIDVASALHTMPSVIRTLVNIYYT	900
Qy	901	ALNKAESNLPSEWDEQTLAENWEAGLSLQOQATLADYTAERLSSVLGNWFLANTQPEGVS	960
Db	901	ALNKAESNLPSEWDEQTLAENWEAGLSLQOQATLADYTAERLSSVLGNWFLANTQPEGVS	960
Qy	961	LHSRDDLYSYFLIDNOVSSAIKTTRLABAIAQIOLYINRALNREPNARADVSTROFFTD	1020
Db	961	LHSRDDLYSYFLIDNOVSSAIKTTRLABAIAQIOLYINRALNREPNARADVSTROFFTD	1020
Qy	1021	WTVNNRYSTWGGVSRVLVYPENYIDPTQRIQOTRMDELLENISOSKLSRDTVEDAFKTY	1080
Db	1021	WTVNNRYSTWGGVSRVLVYPENYIDPTQRIQOTRMDELLENISOSKLSRDTVEDAFKTY	1080
Qy	1081	LTRPETVADIAKVSAHYDNVNSNTGLTWFGQGTRENTPEYYWRNVDISRMQAGELAAW	1140
Db	1081	LTRPETVADIAKVSAHYDNVNSNTGLTWFGQGTRENTPEYYWRNVDISRMQAGELAAW	1140
Qy	1141	KEWKIDTAVNPYKDAIRPVLFRERLHLIWEKEVAKNGTDPVETYDRFTLKLAFLRHD	1200
Db	1141	KEWKIDTAVNPYKDAIRPVLFRERLHLIWEKEVAKNGTDPVETYDRFTLKLAFLRHD	1200
Qy	1201	GSWAPMSYDITTOVEAVTDKPKDTERLALAASGFQGEDTLLVFVYKTKGYSDFGGSNK	1260
Db	1201	GSWAPMSYDITTOVEAVTDKPKDTERLALAASGFQGEDTLLVFVYKTKGYSDFGGSNK	1260
Qy	1261	NVAGMTIYGDGSKFMENTALSRYSQLKNTFDIHTQGNDLVRKASYRPAQDFEVPASLN	1320
Db	1261	NVAGMTIYGDGSKFMENTALSRYSQLKNTFDIHTQGNDLVRKASYRPAQDFEVPASLN	1320
Qy	1321	MSAIGDSSLTWMENGIPOITSKYSSDNLAITLHNAAFTRYDGSNVIRNKQISAMKL	1380
Db	1321	MSAIGDSSLTWMENGIPOITSKYSSDNLAITLHNAAFTRYDGSNVIRNKQISAMKL	1380
Qy	1381	TGVDGSKSYGNAFIANTVKHYGYSIDLGGPITVYNKTKNYIASVQGHLMNADYTRRLIL	1440
Db	1381	TGVDGSKSYGNAFIANTVKHYGYSIDLGGPITVYNKTKNYIASVQGHLMNADYTRRLIL	1440
Qy	1441	TPVENNYARLPEPPFSNTILNTVFTVGSNKTSDFKCSYAVDGNNSQGFIFSSYQSS	1500
Db	1441	TPVENNYARLPEPPFSNTILNTVFTVGSNKTSDFKCSYAVDGNNSQGFIFSSYQSS	1500
Qy	1501	GLWLDITGINTNDIKITWAGSKTHTFTASDHIAASLPANSFPAMPYTFKPLEIDASSLAF	1560
Db	1501	GLWLDITGINTNDIKITWAGSKTHTFTASDHIAASLPANSFPAMPYTFKPLEIDASSLAF	1560
Qy	1561	TNNIAPLDIVFETKAKDGRVLGKIKOTLSVKRVNYPEDILFLRETHSGAQYMLGVYRI	1620
Db	1561	TNNIAPLDIVFETKAKDGRVLGKIKOTLSVKRVNYPEDILFLRETHSGAQYMLGVYRI	1620
Qy	1621	RLNTLLASQLVSRANTGIDTILTMETQRLPEPLGEFFANFVLPKYDPAEHDGRWEFKI	1680
Db	1621	RLNTLLASQLVSRANTGIDTILTMETQRLPEPLGEFFANFVLPKYDPAEHDGRWEFKI	1680
Qy	1681	HIGNVGGNTGRQPYYSGLMISDTSMTLFPYAEGYTMHEGVLGVGYQKITDYNTWES	1740
Db	1681	HIGNVGGNTGRQPYYSGLMISDTSMTLFPYAEGYTMHEGVLGVGYQKITDYNTWES	1740
Qy	1741	AFYFDETKQOQFVLINDADHDSGMTQOQGIKVNKKYKGFNLVSIATGYSAPMDFNASAL	1800
Db	1741	AFYFDETKQOQFVLINDADHDSGMTQOQGIKVNKKYKGFNLVSIATGYSAPMDFNASAL	1800

Qy	1801	YYWELFYTPMWCFORLLQEQFDEATQWIVYNVPAGYIVNGEIAPIWNCRPLEETTS	1860
Db	1801	YYWELFYTPMWCFORLLQEQFDEATQWIVYNVPAGYIVNGEIAPIWNCRPLEETTS	1860
Qy	1861	WNANPLDAIDPDAVAQNDDPMHYKIATFRLLDQLILRGDMAYRELTRDALNEAKWYVRT	1920
Db	1861	WNANPLDAIDPDAVAQNDDPMHYKIATFRLLDQLILRGDMAYRELTRDALNEAKWYVRT	1920
Qy	1921	LELLGDEPDYGSQOAAAPSLSGAASQTVQAAAYOODLTMLGRGGVSKNLRANSILVGLFL	1980
Db	1921	LELLGDEPDYGSQOAAAPSLSGAASQTVQAAAYOODLTMLGRGGVSKNLRANSILVGLFL	1980
Qy	1981	PEYNPALTDYQOTLRLRLFNLRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASQGSASV	2040
Db	1981	PEYNPALTDYQOTLRLRLFNLRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASQGSASV	2040
Qy	2041	LPGLTSLYRFPVMLERLNLVAQLTQFGTSLLSMAEHDDADELTLLLOQGMELATQSI	2100
Db	2041	LPGLTSLYRFPVMLERLNLVAQLTQFGTSLLSMAEHDDADELTLLLOQGMELATQSI	2100
Qy	2101	IQORTVDEVDADIIVLAESRBSAQNRLEKYQOQLYDEIDINHGEORAMSLDDAAAGOSLAGQ	2160
Db	2101	IQORTVDEVDADIIVLAESRBSAQNRLEKYQOQLYDEIDINHGEORAMSLDDAAAGOSLAGQ	2160
Qy	2161	VLSIABGVADILVNVFGLACGSRWGAALRASAVMSLSATASQVSADKISRSEAYRERR	2220
Db	2161	VLSIABGVADILVNVFGLACGSRWGAALRASAVMSLSATASQVSADKISRSEAYRERR	2220
Qy	2221	QEWEIFQDNADGEVQKMDAQLESKIRREAAQMVQYQETQQAHTQAOLELLQKFTNKA	2280
Db	2221	QEWEIFQDNADGEVQKMDAQLESKIRREAAQMVQYQETQQAHTQAOLELLQKFTNKA	2280
Qy	2281	LYSWMRGKLSAIYQFDFLTQSFCLMAQEALRRELITONGVTFFIRGGAWNGTTAGLMAGET	2340
Db	2281	LYSWMRGKLSAIYQFDFLTQSFCLMAQEALRRELITONGVTFFIRGGAWNGTTAGLMAGET	2340
Qy	2341	LLNLAEKWKVLERDERALEVTRVSLAQFYQALSSDNFNLTEKLTQFLREGKGNVGAS	2400
Db	2341	LLNLAEKWKVLERDERALEVTRVSLAQFYQALSSDNFNLTEKLTQFLREGKGNVGAS	2400
Qy	2401	GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG	2460
Db	2401	GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG	2460
Qy	2461	GSIVMPRCGSAIALSHGVNDSGQFMDPNDSDRYLPFEGISVNDGSLTSLSPDATDROKA	2520
Db	2461	GSIVMPRCGSAIALSHGVNDSGQFMDPNDSDRYLPFEGISVNDGSLTSLSPDATDROKA	2520
Qy	2521	LLESLSDIILHIRTIRS	2538
Db	2521	LLESLSDIILHIRTIRS	2538
RESULT 3			
ADR21514			
ID	ADR21514	standard; protein; 2538 AA.	
AC	ADR21514;		
XX			
DT	04-NOV-2004	(first entry)	
XX			
DE	Xenorhabdus strain Xwi cosmid pDAB2097 ORF7 deduced protein SEQ ID NO:34.		
XX	toxin; insect; insecticidal; transgenic; pest control; cosmid.		
OS	Xenorhabdus nematophila.		
XX			
PN	WO2004067727-A2.		
XX			
PD	12-AUG-2004.		
XX			
PF	07-JAN-2004; 2004WO-US000394.		
XX			

PR 21-JAN-2003; 2003US-0441723P.
 XX (DOWC) DOW AGROSCIENCES LLC.
 XX Hey JD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
 XX Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
 XX WPI; 2004-580999/56.
 XX N-PSDB; ADR21499.
 XX Controlling or inhibiting an insect, useful for pest control, comprises
 XX contacting the insect with effective amounts of a Protein A, a Protein B,
 XX and a Protein C.
 XX Claim 1; SEQ ID NO 34; 368pp; English.
 XX The invention relates to a novel method for controlling or inhibiting an
 XX insect comprising contacting the insect with effective amounts of a
 XX Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
 XX is encoded by a naturally occurring gene or has an amino acid sequence
 XX that differs from the product encoded by a naturally occurring gene only
 XX by truncation or by conservative amino acid changes. Protein A is a 230-
 XX 290 kDa toxin complex insect toxin that is derived from a first taxonomic
 XX species, has stand alone insecticidal activity, and has an amino acid
 XX sequence at least 40% identical to a sequence selected from xptA1wi,
 XX xptA2wi, TcdA, TcdA2, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin
 XX complex potentiator having an amino acid sequence at least 40% identical
 XX to a sequence selected from TcdB1, TcdB2, TcdC, xptC1wi, xptC1xb,
 XX xptB1(orfs), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
 XX having an amino acid sequence at least 35% identical to a sequence
 XX selected from TcdC1, TcdC2, TcdC3, TcdC4, TcdC5, xptB1wi, xptC1xb, PptC1
 XX (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic
 XX plant or plant cell that produces a Protein A, a Protein B, and a Protein
 XX C. The method is useful for pest control. The present sequence represents
 XX the deduced protein of ORF7 (xptA2wi) of cosmid insert pBAB2097, obtained
 XX from Xenorhabdus nematophila strain Xwi.
 XX Sequence 2538 AA;
 XX Query Match 100.0%; Score 13043; DB 8; Length 2538;
 XX Best Local Similarity 100.0%; Pred. No. 0;
 XX Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MYSTAVLLNKISPTRDGQMTLADLQYLSFSELKIFDDQLSWGBARHLHYHETIBQKKN 60
 DB 1 MYSTAVLLNKISPTRDGQMTLADLQYLSFSELKIFDDQLSWGBARHLHYHETIBQKKN 60
 QY 61 RLLEARIETFRANPOLSGAIRLGIERSVSRSDYDEMFGARSSFVKPGSVASMPSPAGYLT 120
 DB 61 RLLEARIETFRANPOLSGAIRLGIERSVSRSDYDEMFGARSSFVKPGSVASMPSPAGYLT 120
 QY 121 ELYREAKDLHFSSAYHLDNRPPDLADLTLSQSNMDETLSTLSNELLLEHITRKTGDD 180
 DB 121 ELYREAKDLHFSSAYHLDNRPPDLADLTLSQSNMDETLSTLSNELLLEHITRKTGDD 180
 QY 181 SDALMESLSTYRQAIDTPHQYETIRQVIMTHDSTLSALSRLNPEVMQAGASLAILA 240
 DB 181 SDALMESLSTYRQAIDTPHQYETIRQVIMTHDSTLSALSRLNPEVMQAGASLAILA 240
 QY 241 NISPELYNLTBEITEKNADALFAQNFSENITPENFASQSWIAKYGGLSEVQKYLGM 300
 DB 241 NISPELYNLTBEITEKNADALFAQNFSENITPENFASQSWIAKYGGLSEVQKYLGM 300
 QY 301 ONQYSDSTSAVDNISTGLVNNESKLEAYKITRKTDDYDKNINFDLMYEGNNOFFIR 360
 DB 301 ONQYSDSTSAVDNISTGLVNNESKLEAYKITRKTDDYDKNINFDLMYEGNNOFFIR 360
 QY 361 ANFKVSRREGATLRKNAGPSGIVGSLGGLIANTNFKSNLYLSNISDSEYKNGVKIYAYRY 420
 DB 361 ANFKVSRREGATLRKNAGPSGIVGSLGGLIANTNFKSNLYLSNISDSEYKNGVKIYAYRY 420
 QY 421 TSSTSATNQGGIIFTFESYPLTIFALKLNKAIKRLCLTSGLSNPELOTIVRSNDAQIIND 480

DB 421 TSSTSATNQGGIIFTFESYPLTIFALKLNKAIKRLCLTSGLSNPELOTIVRSNDAQIIND 480
 QY 481 SVLTAKVFTYLYSHRYALSPDDAQVLYNGSVINQVADDDSVSHFNRLFNTPPLKGIKIFAD 540
 DB 481 SVLTAKVFTYLYSHRYALSPDDAQVLYNGSVINQVADDDSVSHFNRLFNTPPLKGIKIFAD 540
 QY 541 GNTVISIDPDEEQSTFARSALMRGIGVNSGELYQJGKLAGVLDQAQNTITLSVVFVSSLYRL 600
 DB 541 GNTVISIDPDEEQSTFARSALMRGIGVNSGELYQJGKLAGVLDQAQNTITLSVVFVSSLYRL 600
 QY 601 TLLARVHQLTVNELCMLYGLSPFNGKTTASLSSGELPRLVILWLYQVTLQWLTAEAITTEAI 660
 DB 601 TLLARVHQLTVNELCMLYGLSPFNGKTTASLSSGELPRLVILWLYQVTLQWLTAEAITTEAI 660
 QY 661 WLLCTPSPGNSISPEISNLNLRPSISEDMAQSHNRELOAEIILAPFIAATLHLASPDMA 720
 DB 661 WLLCTPSPGNSISPEISNLNLRPSISEDMAQSHNRELOAEIILAPFIAATLHLASPDMA 720
 QY 721 RYILLWTDLNLRPGGLDIAGFMVTLVKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780
 DB 721 RYILLWTDLNLRPGGLDIAGFMVTLVKESLNANETTLQVQFCHVMAQLSLSVQTLRLSEAE 780
 QY 781 LSVLVISGFVILGAKNQAPQGHNDITLPSLYRFHQWINGLGNPGSDTLDMLRQOTLTADR 840
 DB 781 LSVLVISGFVILGAKNQAPQGHNDITLPSLYRFHQWINGLGNPGSDTLDMLRQOTLTADR 840
 QY 841 LASVGLDLSMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLNIRYVT 900
 DB 841 LASVGLDLSMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLNIRYVT 900
 QY 901 ALNKAESNLPSEDEWQTLAENWEAGLSLQOQATLADYTAERLSSVLCNWFELANIPEGVS 960
 DB 901 ALNKAESNLPSEDEWQTLAENWEAGLSLQOQATLADYTAERLSSVLCNWFELANIPEGVS 960
 QY 961 LHSRDDLYSYFLIDNQVSSAIIKTRLAELIAGIOLYINRNLNRIEPNARADVSTQFFTD 1020
 DB 961 LHSRDDLYSYFLIDNQVSSAIIKTRLAELIAGIOLYINRNLNRIEPNARADVSTQFFTD 1020
 QY 1021 WTVNRRYSTWGVSRVLYPENYIDPTQRIQOTRMDLLENISOSKLSRDTVEAFKTY 1080
 DB 1021 WTVNRRYSTWGVSRVLYPENYIDPTQRIQOTRMDLLENISOSKLSRDTVEAFKTY 1080
 QY 1081 LTRPETVADLVKVSAYHNDVNSNTGLTWFGQGTRENLPEYWRNVDISRMQAGELAANAW 1140
 DB 1081 LTRPETVADLVKVSAYHNDVNSNTGLTWFGQGTRENLPEYWRNVDISRMQAGELAANAW 1140
 QY 1141 KEWKIIDLAVNPYKDAIRPVIKPRERHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200
 DB 1141 KEWKIIDLAVNPYKDAIRPVIKPRERHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200
 QY 1201 GSWAPMSYDITTOVEAVTDKKPOTERLALAASGFGEDTLLVFFYTKGKSYSDFGGSK 1260
 DB 1201 GSWAPMSYDITTOVEAVTDKKPOTERLALAASGFGEDTLLVFFYTKGKSYSDFGGSK 1260
 QY 1261 NVAGMTIYGDGSKFKMENTALSRYSQLKNTFDIHTQGNDLVRKASYFAQDFEVPASLN 1320
 DB 1261 NVAGMTIYGDGSKFKMENTALSRYSQLKNTFDIHTQGNDLVRKASYFAQDFEVPASLN 1320
 QY 1321 MGSAGDSDSLTWENGNIPQITTSKYSSDNLAITLHNAAFVTRYDGSNGVNRNKOISAMKL 1380
 DB 1321 MGSAGDSDSLTWENGNIPQITTSKYSSDNLAITLHNAAFVTRYDGSNGVNRNKOISAMKL 1380
 QY 1381 TGVDGSKSYGNAFIANTVKGHYGSDLGGPITVYNTKNTNYIASVOGHLNADYTRRLIL 1440
 DB 1381 TGVDGSKSYGNAFIANTVKGHYGSDLGGPITVYNTKNTNYIASVOGHLNADYTRRLIL 1440
 QY 1441 TPVENNYIARLPPEPPSPNTILNTVFTVGSNKTSDFKCSYAVDGNNSGFOIFSSYQSS 1500
 DB 1441 TPVENNYIARLPPEPPSPNTILNTVFTVGSNKTSDFKCSYAVDGNNSGFOIFSSYQSS 1500
 QY 1501 GWLDDTGINNTDIIKITYWAGSKHTTPTASDHIAASLPANSPDAMPYTPKPLEIDASSLAF 1560
 DB 1501 GWLDDTGINNTDIIKITYWAGSKHTTPTASDHIAASLPANSPDAMPYTPKPLEIDASSLAF 1560

QY 1561 TNNIAPLDIVPETKADGRVLGKIQTLSVKRVNYPNPDILFLRETHSGAOTMOLGVVRI 1620
Db 1561 TNNIAPLDIVPETKADGRVLGKIQTLSVKRVNYPNPDILFLRETHSGAOTMOLGVVRI 1620
QY 1621 RLNTLLASQLVSRANTGIDTILTMETQRLPEPLGEGFFANFVLPKYDPAEHDGRWFKI 1680
Db 1621 RLNTLLASQLVSRANTGIDTILTMETQRLPEPLGEGFFANFVLPKYDPAEHDGRWFKI 1680
QY 1681 HIGNVGGNTGQPYYSGLMDSSTSETMTLFPYABGYTHGVRGLGVGYKITYDNTWES 1740
Db 1681 HIGNVGGNTGQPYYSGLMDSSTSETMTLFPYABGYTHGVRGLGVGYKITYDNTWES 1740
QY 1741 AFFYFDETKQOQVFLINDADHDSGMTQOGIVKNIKKYKGLNVSATGYSAPMDFNFSAL 1800
Db 1741 AFFYFDETKQOQVFLINDADHDSGMTQOGIVKNIKKYKGLNVSATGYSAPMDFNFSAL 1800
QY 1801 YWELFYTTWMCFORLLQEQFDEATQWVNVNYPAGYIYNGEITAPWIMNCRPLEETTS 1860
Db 1801 YWELFYTTWMCFORLLQEQFDEATQWVNVNYPAGYIYNGEITAPWIMNCRPLEETTS 1860
QY 1861 WNAFLDAIDPDVAQNPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920
Db 1861 WNAFLDAIDPDVAQNPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920
QY 1921 LELIGDEPEDYGSQOQWAPSLSGAASQTVQAAQODLTWLGKGGVSKNLTANSILVGLFL 1980
Db 1921 LELIGDEPEDYGSQOQWAPSLSGAASQTVQAAQODLTWLGKGGVSKNLTANSILVGLFL 1980
QY 1981 PEYNPALTDYWTLLRLFLNLRHNLSDIQPLSLAIYAEPTDPKALLTSMVQASQGSVAV 2040
Db 1981 PEYNPALTDYWTLLRLFLNLRHNLSDIQPLSLAIYAEPTDPKALLTSMVQASQGSVAV 2040
QY 2041 LPGTLSLYRFPVWMLERTNLVAQLTQFGTSLLSMAEHDDADELTLLLLQQQWELATQSI 2100
Db 2041 LPGTLSLYRFPVWMLERTNLVAQLTQFGTSLLSMAEHDDADELTLLLLQQQWELATQSI 2100
QY 2101 IQQRTVDEVDADIAVLASRRSAONREKYQOQYDEIDNHGEORAMSLDDAAAGOSLAGO 2160
Db 2101 IQQRTVDEVDADIAVLASRRSAONREKYQOQYDEIDNHGEORAMSLDDAAAGOSLAGO 2160
QY 2161 VLSIAEGVADLVPNVFLGACGSRWGAAALRASAVWSLSATASQYSADKISRSEAYRRRR 2220
Db 2161 VLSIAEGVADLVPNVFLGACGSRWGAAALRASAVWSLSATASQYSADKISRSEAYRRRR 2220
QY 2221 QEWETQRDNADGEVKQMDAQLESKIRREAAQMVVEYQETOQAHTQAQLELLQKFTNKA 2280
Db 2221 QEWETQRDNADGEVKQMDAQLESKIRREAAQMVVEYQETOQAHTQAQLELLQKFTNKA 2280
QY 2281 LYSWMRGKLSAIYYQFPLDTQSFCMAQEARRELTNDGVTFIRGAWNGTTAGLMAGET 2340
Db 2281 LYSWMRGKLSAIYYQFPLDTQSFCMAQEARRELTNDGVTFIRGAWNGTTAGLMAGET 2340
QY 2341 LLLNLAEMKWLDERALEVTRTVSLAQFYQALSSDNFNLTEKLTQFLREGKNGV GAS 2400
Db 2341 LLLNLAEMKWLDERALEVTRTVSLAQFYQALSSDNFNLTEKLTQFLREGKNGV GAS 2400
QY 2401 GNEKLSNRQIEASVRLSDLKIFSDYPSLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG 2460
Db 2401 GNEKLSNRQIEASVRLSDLKIFSDYPSLGNTRQLKQVSVTLPALVGPYEDIRAVLNYG 2460
QY 2461 GSIWVPRGCSAIALSHGVNDGQFMDLDFNDSRYLPPEGISVNDGSLTSPDATDRQKA 2520
Db 2461 GSIWVPRGCSAIALSHGVNDGQFMDLDFNDSRYLPPEGISVNDGSLTSPDATDRQKA 2520
QY 2521 LLESLSDIILHIRTYS 2538
Db 2521 LLESLSDIILHIRTYS 2538

RESULT 4
ADR21500
ID ADR21500 standard; protein; 2538 AA.

XX ADR21500;
XX 04-NOV-2004 (first entry)
XX Xenorhabdus strain Xwi cosmid pDAB2097 ORF7 deduced protein SEQ ID NO:20.
XX toxin; insect; insecticidal; transgenic; pest control; cosmid.
XX Xenorhabdus nematophila.
XX WO2004067727-A2.
XX 12-AUG-2004.
XX 07-JAN-2004; 2004WO-US000394.
XX 21-JAN-2003; 2003US-0441723P.
XX (DOWC) DOW AGROSCIENCES LLC.
XX Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX Ni W, Zhu B, Merlo DU, Apel-Birkhold PC;
XX WPI; 2004-580999/56.
XX N-PSDB; ADR21499.
XX Controlling or inhibiting an insect, useful for pest control, comprises
XX contacting the insect with effective amounts of a Protein A, a Protein B,
XX and a Protein C.
XX Claim 12; SEQ ID NO 20; 368pp; English.
XX The invention relates to a novel method for controlling or inhibiting an
XX insect comprising contacting the insect with effective amounts of a
XX Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
XX is encoded by a naturally occurring gene or has an amino acid sequence
XX that differs from the product encoded by a naturally occurring gene only
XX by truncation or by conservative amino acid changes. Protein A is a 230-
XX 290 kDa toxin complex insect toxin that is derived from a first taxonomic
XX species, has stand alone insecticidal activity, and has an amino acid
XX sequence at least 40% identical to a sequence selected from XprAlwi,
XX XpA2wi, Tcda, TcdA2, TcdA4, and TcBA. Protein B is a 130-180 kDa toxin
XX complex potentiator having an amino acid sequence at least 40% identical
XX to a sequence selected from TcdB1, TcdB2, TcAc, XptCiwi, XptB1xb,
XX PpC1(orfs), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
XX having an amino acid sequence at least 35% identical to a sequence
XX selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1wi, XptC1xb, PpC1
XX (orf 6 long), PpC1 (orf 6 short), and SepC. Also claimed is a transgenic
XX plant or plant cell that produces a Protein A, a Protein B, and a Protein
XX C. The method is useful for pest control. The present sequence represents
XX the deduced protein of ORF7 (XpA2wi) of cosmid insert pDAB2097, obtained
XX from Xenorhabdus nematophila strain Xwi.

Sequence 2538 AA;

Query Match 100.0%; Score 13043; DB 8; Length 2538;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSTAVILNKISPRDQGTMTLADQLVLSFSELAKIFDDQLSWGAEARHLHYETIQKKN 60
Db 1 MYSTAVILNKISPRDQGTMTLADQLVLSFSELAKIFDDQLSWGAEARHLHYETIQKKN 60
QY 61 RLLEARIPTANPOLSGAIRLGIERSVSRSDYEMFGARSSSVFKPGSVASMFSPAGYLT 120
Db 61 RLLEARIPTANPOLSGAIRLGIERSVSRSDYEMFGARSSSVFKPGSVASMFSPAGYLT 120
QY 121 ELYREAKDLHFSSAYHLNDRRPLADLTLSQSNMDTEISTLTLSNELLLEHITRKTCGD 180
Db 121 ELYREAKDLHFSSAYHLNDRRPLADLTLSQSNMDTEISTLTLSNELLLEHITRKTCGD 180
QY 181 SDALMESLSTVRQAIDTPYHQPYETIRQVIMTHDSTLSALSRNPVEMQAGSALLAILA 240

Db 181 SDALMESLTYRQADTDPHQYETIRQVIMTHDSTLSALSBNPEVMQAGASLAILA 240
QY 241 NISPELYNLTBEITEKADALFAQNFSENIIPENFASQSWIAKYVGLSEVQKYLGM 300
Db 241 NISPELYNLTBEITEKADALFAQNFSENIIPENFASQSWIAKYVGLSEVQKYLGM 300
QY 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKTRVKTDDYDXNINYPDLMEGNNQFFIR 360
Db 301 QNGYSDSTSAVDNISTGLVNNESKLEAYKTRVKTDDYDXNINYPDLMEGNNQFFIR 360
QY 361 ANFKYSRBFATLRKNAGSGIVGSLGPLIANTNPKSNYLSNISDSSEYKNGVKIYAYRY 420
Db 361 ANFKYSRBFATLRKNAGSGIVGSLGPLIANTNPKSNYLSNISDSSEYKNGVKIYAYRY 420
QY 421 TSSTSATNQGCGIFTFESYPLTFIPALKLNKARLCLTSGLSNRELQTVRSNDAQIIND 480
Db 421 TSSTSATNQGCGIFTFESYPLTFIPALKLNKARLCLTSGLSNRELQTVRSNDAQIIND 480
QY 481 SVLTQVFTYTLFYSHYALSFDDAQVLNGSVINQYADDSVSHFNRLFPNTPPLKGIPEAD 540
Db 481 SVLTQVFTYTLFYSHYALSFDDAQVLNGSVINQYADDSVSHFNRLFPNTPPLKGIPEAD 540
QY 541 GNTVSDIDPDEOSTFARSALMRLGVLNGSELQYOLKLAGVLDQNTIILSVFVSSLYRL 600
Db 541 GNTVSDIDPDEOSTFARSALMRLGVLNGSELQYOLKLAGVLDQNTIILSVFVSSLYRL 600
QY 601 TLLARVHQLTVNELCWLGLSPFNKTTASLSSGELPRVILWYQVOWLWTEABETTRAI 660
Db 601 TLLARVHQLTVNELCWLGLSPFNKTTASLSSGELPRVILWYQVOWLWTEABETTRAI 660
QY 661 WLLCTPEFSGNISPEISLNNLRPSISBDMQASHNRELQAEILAPFIAATLHLASPDMA 720
Db 661 WLLCTPEFSGNISPEISLNNLRPSISBDMQASHNRELQAEILAPFIAATLHLASPDMA 720
QY 721 RYLLWTONLRPGGDIAGFMTLVLKESLNANETTLQVQFCHMAQLSLSVQTLSEAE 780
Db 721 RYLLWTONLRPGGDIAGFMTLVLKESLNANETTLQVQFCHMAQLSLSVQTLSEAE 780
QY 781 LSVLVISQFAVLGAKNQAGQHNIDTFLSLRFHQWINGLNGPSGDTLDMRLQOQTLTADR 840
Db 781 LSVLVISQFAVLGAKNQAGQHNIDTFLSLRFHQWINGLNGPSGDTLDMRLQOQTLTADR 840
QY 841 LASVGLDISMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900
Db 841 LASVGLDISMVTQAMVSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNIYVT 900
QY 901 ALNKAESNLPSEWQOTLAENWEAGLSTQOAGTLADYTAERLSSVLCNWFLANTIOPEGVS 960
Db 901 ALNKAESNLPSEWQOTLAENWEAGLSTQOAGTLADYTAERLSSVLCNWFLANTIOPEGVS 960
QY 961 LHSRDDLYSYFLIDNQVSSAIKTRFLAELAGIQLYINRNLNRIIPNARADVSTROQFTD 1020
Db 961 LHSRDDLYSYFLIDNQVSSAIKTRFLAELAGIQLYINRNLNRIIPNARADVSTROQFTD 1020
QY 1021 WTVNNRSTWGVSRVLVYPENYIDPTQRIQGTMMDELLNENISQSLSRDTVEDAFKTY 1080
Db 1021 WTVNNRSTWGVSRVLVYPENYIDPTQRIQGTMMDELLNENISQSLSRDTVEDAFKTY 1080
QY 1081 LTRPETVADLKVVSAYHDNVNSNTGLTWFGVGTRENLPYYWRVNDISRMQAGELAAANW 1140
Db 1081 LTRPETVADLKVVSAYHDNVNSNTGLTWFGVGTRENLPYYWRVNDISRMQAGELAAANW 1140
QY 1141 KEWKIDTAVNPKDAIRPVIFRERLHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200
Db 1141 KEWKIDTAVNPKDAIRPVIFRERLHLIWEKEEVAKNGTDPVETYDRFTLKLAFLRHD 1200
QY 1201 GWSAPMSYDITTOVEAVTDKPDTERLALASGFGQEDTLLVYFKTKGYSDFGGSNK 1260
Db 1201 GWSAPMSYDITTOVEAVTDKPDTERLALASGFGQEDTLLVYFKTKGYSDFGGSNK 1260
QY 1261 NVAGHTIYDGSFKQKMENTALSRYSQLKNTFPIHTQGNLDLVRKASYRPAQDFEVPASLN 1320

Db 1261 NVAGHTIYDGSFKQKMENTALSRYSQLKNTFPIHTQGNLDLVRKASYRPAQDFEVPASLN 1320
QY 1321 MGSAGDDSLTWMENGINPQITSKYSSDNLAITLHNAAFTRVYDGSNGVIRNKQISAMKL 1380
Db 1321 MGSAGDDSLTWMENGINPQITSKYSSDNLAITLHNAAFTRVYDGSNGVIRNKQISAMKL 1380
QY 1381 TGVDGKSQYGNAPIANTVKGHYGYSDLGGPITVYNKTKNYIASVOGHLNADYTRRLTL 1440
Db 1381 TGVDGKSQYGNAPIANTVKGHYGYSDLGGPITVYNKTKNYIASVOGHLNADYTRRLTL 1440
QY 1441 TPVENNYARLPEPPSPNTILANTVFTVGSNKTSDFPKCSYAVDGNNSQGFQIFSSYQSS 1500
Db 1441 TPVENNYARLPEPPSPNTILANTVFTVGSNKTSDFPKCSYAVDGNNSQGFQIFSSYQSS 1500
QY 1501 GMLDITDGINNTDIKITWAGSKTHFTTASDHIAISLPANSPDAMPYTPKPLEIDASSLAF 1560
Db 1501 GMLDITDGINNTDIKITWAGSKTHFTTASDHIAISLPANSPDAMPYTPKPLEIDASSLAF 1560
QY 1561 TNNIAPLDIVPETKAKOORVLGKIQTLSVKRVNTPEDILPLRETHSGAQYMLQGVYRI 1620
Db 1561 TNNIAPLDIVPETKAKOORVLGKIQTLSVKRVNTPEDILPLRETHSGAQYMLQGVYRI 1620
QY 1621 RLNTLLASQLSVRANTGIDTILTMETQRLPEPPLGEGFPANFVLKYPDAEHGDERWFKI 1680
Db 1621 RLNTLLASQLSVRANTGIDTILTMETQRLPEPPLGEGFPANFVLKYPDAEHGDERWFKI 1680
QY 1681 HIGNVGNTGRQPYYSGLMSTSETSMTLFPVYAEYTMHGEVRLGVGYQKITTYDNTWBS 1740
Db 1681 HIGNVGNTGRQPYYSGLMSTSETSMTLFPVYAEYTMHGEVRLGVGYQKITTYDNTWBS 1740
QY 1741 AFFPDETQKQFVLINDADHDSGMTQOQIVKNIKKYKGLNVIATGYSAPMDFNSASAL 1800
Db 1741 AFFPDETQKQFVLINDADHDSGMTQOQIVKNIKKYKGLNVIATGYSAPMDFNSASAL 1800
QY 1801 YWELFYTPMCMFORLLOEQFDEATOWINVYNPAGYVINGETAPMWNCRPLEETTS 1860
Db 1801 YWELFYTPMCMFORLLOEQFDEATOWINVYNPAGYVINGETAPMWNCRPLEETTS 1860
QY 1861 WNANPLDAIDPDVAQNDDPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920
Db 1861 WNANPLDAIDPDVAQNDDPMHYKIATFMRLLDQLILRGDMAYRELTRDALNEAKWYVRT 1920
QY 1921 LELIGDEPEDYGSQWAAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSVLGFL 1980
Db 1921 LELIGDEPEDYGSQWAAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSVLGFL 1980
QY 1981 PEYNPALTDYVQTLRLRLFNLRHNLSDGQPLSLAIYAEPTDPPKALLTSMVQASOGGSV 2040
Db 1981 PEYNPALTDYVQTLRLRLFNLRHNLSDGQPLSLAIYAEPTDPPKALLTSMVQASOGGSV 2040
QY 2041 LPGTSLSYRFPVMLERTNLVAQLTQFCTSLLSMAEHDDADELTLLLLQOQMELATQSIR 2100
Db 2041 LPGTSLSYRFPVMLERTNLVAQLTQFCTSLLSMAEHDDADELTLLLLQOQMELATQSIR 2100
QY 2101 IQORTVDEVDADIIVLAEBSRSANRLSKYQOOLYDEIDINHGEOQAMSLDAAAGOSLAGQ 2160
Db 2101 IQORTVDEVDADIIVLAEBSRSANRLSKYQOOLYDEIDINHGEOQAMSLDAAAGOSLAGQ 2160
QY 2161 VLSIABGVADIVPNVFGIACGSRWGAALRASAVMSISATASQYSAADKISRSEAYRRRR 2220
Db 2161 VLSIABGVADIVPNVFGIACGSRWGAALRASAVMSISATASQYSAADKISRSEAYRRRR 2220
QY 2221 QEWIQRDNADGEVKQMDAQLESKIRREAAQMOVYQETOQAHTQAQLELLQRKFTNKA 2280
Db 2221 QEWIQRDNADGEVKQMDAQLESKIRREAAQMOVYQETOQAHTQAQLELLQRKFTNKA 2280
QY 2281 LYSWMRGKLSIYYQFPDLTOSFCLMAQEALRRELTDNGVTFFIRGGAWNGTTAGLMAGET 2340
Db 2281 LYSWMRGKLSIYYQFPDLTOSFCLMAQEALRRELTDNGVTFFIRGGAWNGTTAGLMAGET 2340
QY 2341 LLLNLAEMEKVLRDERALSVTRTVSLAQFQALSSDNFNLTEKLTQFLREGKGNVCAS 2400
Db 2341 LLLNLAEMEKVLRDERALSVTRTVSLAQFQALSSDNFNLTEKLTQFLREGKGNVCAS 2400

1072 TVEDAFKTYLTRPETVADLKVVSAHYNDVNSNTGLTWFGVGTRENLPEYIYWRNVDIRMQ 1131
1061 TVEDSFKNLTAFEDVANLQVISHYDSINVNEGLTYLIGYSQTEPRYIYWRNVDHQKQ 1120
1132 AGELANAWKWTKIDTAVNPYKDAIRPVI PRERHLIINWEKEEVAKNCTDPVEYIDRPT 1191
1121 HQGFANANWGWKKIEIPINNVQWENIRPVIYKSRLLYLLWLEQKLNESDEGKIDITDI 1180
1192 LKLAFLRHDSGSAFWSYDITQVAVTDKQDPTERLALAAAGFOCEDTLLVFVYKTKGS 1251
1181 LKLSHRYDGSWSPFNVTDKIENLNKAS---IGNYCSDEYKDVIIYVPEKKDN 1237
1252 YSDFGSKNKNVAGTIYDGSFKKMENTALSRYSQLNTFDIIHTQGNLDVRKASVRFPAQ 1311
1238 YS--FNSLPAREGMTINPDMTSLTENDLD--AIVKSTLSBLDTRTE---YKVNQFAT 1290
1312 DP-----EVPASLNWGSAG--DDSLTVMNGNI-----PQITSKYSSDN----- 1349
1291 DYLAEBYKESITTKNKLASFNGIFDLSYISPGNGHINLTFFNPSMEINFSGNIYNDVKY 1350
1350 -LAIITLHNAFTVRDGSNGVIRNKKQISAMKLTGVGDKSOYGNAPFIANTVKHYGYSDL 1408
1351 LLSWVEDETVILFDVRHDEM-----GKEE-----EVHYG----- 1382
1409 GGPITVYNTKNYIASVOGHLNADYTRRL-----ILTPVENNYIARLPF 1454
1383 -----TLDFIISID--LKNAEYFRVLMHLRTEKIPRSEKSEIGVINGYDYESNDAEF 1431
1455 PPSPTIILN-----TVFTVGSKNTSDFK-----KCSYAVDGNNS 1488
1432 KLDITNVLVDKONTGVMTHTICESPTNDVSIIINMGNIALFLREDFCVVYLCISINTD---- 1487
1489 QGQFIFSSYQSSGLDIDGIIINNTDKITVMAGSKTHT--FTASDHIASLPANSPAMPYT 1547
1488 ---IKTASSMIEQ-----IODKNISFLKNGSDILVELNADHVASKPSESHPMYD 1537
1548 FKPEIDASSLAFTNNIAPLDIVFETKAKDGRVLG-KIKQTLVSRVKNVYNPEDIILFLRBT 1606
1538 FNQVKVDIEGYDI-----PLVSEFIKQPDGGYNDIVIESPIHIKLSKDTSNVLSHKM 1592
1607 HSGAQMLQVTRIRLNTLLASQLSVRANTGIDTILTMETQRLPEPLGEGFPANFLPK 1666
1593 PSQTYMQIGPYTRTLNTLFSRKLAFRANIGIDNVLMSMETQNLPEPLGEGFYATFKLPP 1652
1667 YDPAEHGDERWFKIHGNVGGNTGOPYYSGLMSDTSMTLFPYABGYMHGVRIG 1726
1653 YNKEEHGDERWFKIHGNIDGNSAKOPYYEGMLSDI-ETITVTFVPYAKGYIRREGVRLG 1711
1727 VGYQKITDNTWESAFYPFDETKQOFVLINDADHDSGMTQOGIVKNIKKYKGLNVSAT 1786
1712 VGYKKIIVDKSWESAFYFDETKNQFIFINDADHDSGMTQOGIVKNIKKYKGFHVVWK 1771
1787 GYSAPMDFNSASALTYWELFYTPPMCFORLLOEQFDEATOWINYVNPAGIYNGEIA 1846
1772 NNTPEMDFNGANAIYFWEFYTPPMVFORLLQEQNFTESTRLRYIWNYPAGYSVQGENQ 1831
1847 PWIWNCRPLEETSNANPLDAIDPDVAQNDPMHYKIAFMRLDOLLRLCRDMAYRELT 1906
1832 DTYWNVRLPEEDTSNANPLDSVDPDAVAQNDPMHYKIAFMRLDOLLRLTRGDSAYRQLE 1891
1907 RDALNEAKWYVYRTLELLGDEPEDYGSQQWAAAPSLSGAASQTVQAAQQDLTMLRGGVYS 1966
1892 RDTLNEAKWYVYQALTLGDEYFSLDNDWSEPRLEEAASQTRHHYQHKMLQLQRAAL 1951
1967 KNLRTANSVLGFLBEYNPALTDYQTLRLRLFNLRHNLSDIGQPLSLAIYAEPTDPKAL 2026
1952 PTKRTANSITLFLPQINKKLOQYQTLTQRLYNLRHNLTDIGQPLSLSLYATPADPSML 2011
2027 LTSWQASQGSVAFLPGTILSLRYFVPMLETRNLVAQLTQFGTSLLSMAEHDDADELTTL 2086
2012 LSAAITASQGGDLPHAVPMYRFPVILENAKWSQLIQFGNTLLSITERQDAALABI 2071

2087 LLOQWHELATQSIRIQORTVDEVDADIIVLAESRRSAQNRLKEKYQOOLYDEIDINHGEORAM 2146
2072 LQTOGSELALQSKMCKQVMAEIDKALQSRHGAOSRFDNFNTLYDEDYNAGEKQAM 2131
2147 SLLDAAAGSLAGQVLSTAEGVADLVNPNVFLGACGSGRWGAALRASAVMSLSATASQYS 2206
2132 DLYLSSSVLSTSGTALHMAAADLVNPNYIGFVAGSGRFGALFNASATGIESASATRA 2191
2207 ADKISRSBAYRRRRQWEIQRDNADGEVKQMDAQLESKIRREAAQMVVEYQETOQAHTQ 2266
2192 ADKISQSEIYRRRRQWEIQRNNAEAIKQIDAQLATLAVRREAAVLQKNYLETOQAQTQ 2251
2267 AQLELLQKFTNKALYSWNRGKLSAIYQFPDLTOSFCLMAQEALRRELTDNGVTPIRG 2326
2252 AQLAFLQSKFNAALYNLRLGRLSAIYQFYDYLAVSLCLMAEQTYOYELNNAAAHFKPG 2311
2327 ANWGTTAGLMAGETLLNLAEMEKVWLERDERALEVTRTVSLAQFYQALSSDNFNLTCKL 2386
2312 AWHGYTAGLAGETLMLNLAQMEKSYLEKDERALEVTRTVSLAEVYAGLTENSFILKDKV 2371
2387 TOFLREGKNGVASCNELKLSNRQIEASVRLSDLKI FSDYPESLGNTRQLKQVSVTLPAL 2446
2372 TELVNAGEGSGAGTTLNGLNVEGTQLQASLKLSLDNIATDYPDGLGNTRRIKQISVTLPAL 2431
2447 VCPYEDIRAVLNYGSGIWMPCSCSAIALSHGVNDSGQFMDFNDSRYLPEGISVNDSGS 2506
2432 LGPYQDVRAILSYGSGTWMPCSKAIAISHGVNDSGQFMDFNDAKYLPEGLPVADTGT 2491
2507 LTLSPDADTOKALLESLSDIILHRYTIRS 2538
2492 LTLSPFGISGKNSLUSLSDIILHRYTIRS 2523

RESULT 6

ADR21494
ID ADR21494 standard; protein; 2523 AA.
XX AC ADR21494;
XX DT 04-NOV-2004 (first entry)
XX DE Xenorhabdus strain Xwi cosmid pDAB2097 ORF4 deduced protein SEQ ID NO:14.
XX KW toxin; insect; insecticidal; transgenic; pest control; cosmid.
XX OS Xenorhabdus nematophila.
XX PN WO2004067727-A2.
XX PD 12-AUG-2004.
XX PF 07-JAN-2004; 2004WO-US000394.
XX PR 21-JAN-2003; 2003US-0441723P.
XX PA (DOWC) DOW AGROSCIENCES LLC.
XX PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX NI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
XX WPI; 2004-580999/56.
XX DR N-PSDB; ADR21493.
XX PT Controlling or inhibiting an insect, useful for pest control, comprises
XX FT contacting the insect with effective amounts of a Protein A, a Protein B,
XX and a Protein C.
XX PS Claim 1; SEQ ID NO 14; 368pp; English.

CC The invention relates to a novel method for controlling or inhibiting an
CC insect comprising contacting the insect with effective amounts of a
CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
CC is encoded by a naturally occurring gene or has an amino acid sequence

CC that differs from the product encoded by a naturally occurring gene only
 CC by truncation or by conservative amino acid changes. Protein A is a 230-
 CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic
 CC species, has stand alone insecticidal activity, and has an amino acid
 CC sequence at least 40% identical to a sequence selected from xptA2w1,
 CC xptA2w1, TcdA, TcdA2, and TcdB. Protein B is a 130-180 kDa toxin
 CC complex potentiator having an amino acid sequence at least 40% identical
 CC to a sequence selected from TcdB1, TcdB2, TcdC, xptC1w1, xptB1xb,
 CC xptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
 CC having an amino acid sequence at least 35% identical to a sequence
 CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, xptC1w1, xptC1xb, PptC1
 CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic
 CC plant or plant cell that produces a Protein A, a Protein B, and a Protein
 CC C. The method is useful for pest control. The present sequence represents
 CC the deduced protein of ORF4 (xptA2w1) of cosmid insert pDAB2097, obtained
 CC from Xenorhabdus nematophila strain Xw1.
 XX
 SQ

Sequence 2523 AA;

Query Match	42.1%;	Score	5496.5;	DB 8;	Length	2523;
Best Local Similarity	44.7%;	Pred. No. 0;				
Matches 1168;	Conservative	436;	Mismatches	833;	Indels	175;
Gaps	36;					
QY	7	LLNKISPRDGTMTLADQLVLSSEKRLKFDQDLSWGEARHLVHETIEQKKNRLLEAR	66			
DB	7	LLDKINRKSGLTLLTWISFMSFEPHRTSGTLTWRETDLYQQAQESKQNKLELR	66			
QY	67	IFTRANPOLGAIRGIERDSVRSYDEMFGARSSSVKPGSVASMPSPAGLYTLYREA	126			
DB	67	ILSRANPOLANTNLNITPSTLNNISWYFGRHRFVKPGSIASIFSPAAYLTLYREA	126			
QY	127	KDLHSSAYHLNRRPOLADLTLSQNMWTEISTLSNELLLEHTRKTGGSDALWE	186			
DB	127	KDFHPDINSOYHLNKRPRDIASLALQNNDWDEISTLSNELLNHIOTLEKTDYGVNK	186			
QY	187	SLSYRQALDTPYHOPYETIROVIMTHDSTLSALRNPEVWQAGASLLAILANISPEL	246			
DB	187	MLSTYRQGTMPYHLPYESARQAILLODKNLTAFTSNTDVAELMDPTSLIAKTDISEL	246			
QY	247	YNILTEETERKADALFQNP--SENITPENFASQSWIAKYTGLEUSEVKYLGMLQNGYS	305			
DB	247	YQILVEEITPENSTELMKNFCTDVL--IFKSYASLARLYTDLSYDELSLFVN--LSFGKK	303			
QY	306	DSTSAVDNISGLVNNESKLEAYKTRVKTDDYDKINFDLMVYEGNNOFFIRANFKV	365			
DB	304	NTNOQYKNEQLITLVNGDNDTATARKTRKDFYDHLNVAELIPIKENEF--KINFSV	361			
QY	366	SREFCATL--RKNAGPSGIVGSLSGPLIANTNFKSNLSNISDSEYKNGVKIYARYTS	422			
DB	362	KTEPDHLDPRLONGKEYIYQDKNFVPIANTHY--SIPIKLTTEQITNGITLRLWRVKP	419			
QY	423	STSATNOGGIPTFESYPLTIPALKNKAIKRLCTLSGLSPNELQTVIRSDNAQIINDSV	482			
DB	420	NPSDAINANAYFKWMEFGDIFLLKLNKAIKRLKATGISPEDIMQVIESIYDDLIDSNV	479			
QY	483	LTKVFTYLYSHRYALSFDPAQVINGSVINOYADDDSHENRLENTPLKGIPEADGN	542			
DB	480	LKGLFYQYMOHYNISVDALVCHSDISQYSTKQPSHFITLNTPLNGQESADNT	539			
QY	543	TVSIDIPEEQSTFARSALMRGLGVNNGELYQLGKLAGVLDAQNTITLSVFVSIISLYRLTL	602			
DB	540	KLDLTFGSKNHFYLGIMKGAFRVNDTLYTLWLKANG--GTPNPFMCSENLSLYRVL	598			
QY	603	LARVHOLTVELCMLYGLSPFNKKTASLSGSELPRVLVILVQVQWLTTEARITTEALWL	662			
DB	599	LADIHHLTVNELSMLLSPYVNTKIALFSDTALTQLISFLPQCTQWLTQKWSVDVFL	658			
QY	663	LCTPFSGNISPEISNLLNLRPSISEDMAQSHNELQAEILAPTAATLHLASPDMARY	722			
DB	659	MTIDNYSVTLPDIENLITLTSNGUS--TUSLGDDELIRA--AAPLIAASIQMSAKTAE	715			
QY	723	ILLWTDLNRPGLDIAGFTLVLKESLNANETTQLVQFCHVNAQLSLSYQTLRLSEALS	782			

DB	716	ILLWINKIQOGLTFDDFMIAANRDRSENETSNMVAFCQVLGQLSLIVRNLGISENELT	775			
QY	783	VLVISGFAVLGAKNOP-----AGQHNIDTFLSYRFHOWINGLGNPGSDTIDMLRQQ	834			
DB	776	LLV-----TKPEKFOSETTALQHDLPFLOALTRFHAVIMRCGSYATEILTALELG	825			
QY	835	TLTADRLASVGLDISMVTQAM--VSAGVNOQCQMDINTVLOWIDVASALHTWPSVIRT	892			
DB	826	ALTAELQALVALKFDQAVVTQALQQTGLGVNTFTNWRITDVTLOWLDAVAATLGITPDGVA	885			
QY	893	LVNRYVTALNKAESNLPSPDWEQTLAENWEAGLSQQAQTLADYTAERLSSVLNWFELA	952			
DB	886	LILKLYI--GEPETPMTFDDWAQASTLLQAGLNSQOQDQALQAWLDEATTAAASYIK	942			
QY	953	NIQPEGVLSHRDDLVSFLIDNOVSSAIKTRTLAEAIAGIQLYINRNLNRIENPARADV	1012			
DB	943	NSAPQ--QIKSRDELYSYLLIDNQVSAQVKTTRVAEALIASIQLVYRNLANNVEGKVPK	1000			
QY	1013	STROFFTDW--TVNNRYSTWGGVSLVYPENYIDPTQRIQGTMRMDELLENISOSKLSRD	1071			
DB	1001	KTROFFCDWETYNRRYSTWAGVSELAVYPENYIDPTIRIGQTGMNLLQOLSQSLNID	1060			
QY	1072	TVEDAFKTYLTRPETVADLKVVSAYHDNVSNTGLTWFGQGTRENLPYKYHNVDISRMQ	1131			
DB	1061	TVEDSFKNLYTAFEDVANLQVISGHDNSINVEGLTYLIGYSQTEPREYIYWRNVDHQKCQ	1120			
QY	1132	AGELAAANAKWETKIDTAVNPKDAIRPVIPEERLHLIWEKEEVAKNGTDPVETYDRFT	1191			
DB	1121	HGQFAANANGWKKIEIPINWQENIRPVIYKSKLYLLWLQEKELKNESEDKIDITDYI	1180			
QY	1192	LKLAFLRHGDSWSPWSYDITTOVEAVTDKPKDTERLALAAASGFGEDTLVVFYKTKGS	1251			
DB	1181	LKLSHIRYDGSWSFNFNVTDKIENLINKAS--IGMYCSSDYEKDVIIVFHEKKDN	1237			
QY	1252	YSDFGGSKNVAGMTIYGDGSKFKNMENTALSRYSQLKNTFDI IHTQGNDLVRKASYRAQ	1311			
DB	1238	YS--FNSLPAREGTMTPDMLTSLITENDLD--AIVKSTLSELDTRE--YKVNQAPAT	1290			
QY	1312	DP-----EVPASLNMGSAIG--DDSLTVMNGNI-----PQITSKYSSDN-----	1349			
DB	1291	DYLAYPEKESITTKKLKASFTGNI FDL SVI SPNGHINLTFPNMEINFSKGNINYDEVKY	1350			
QY	1350	-LAITLHNAAFTRYDGSQNVIRNKQISAMKLTGVGKSGOYGNAPFIANTVKHGGYSDL	1408			
DB	1351	LLSMVEDETILFDYDRHDEML-----GKEE-----EVFHYG-----	1382			
QY	1409	GGPITVYNTKTYIASVOGHLNADYTRRL-----ILTPVENNYIARLPEF	1454			
DB	1383	-----TLDFIISID--LKNAEYFRVLHMLRTEKIPRKSEIGVGINVDYESNDAEF	1431			
QY	1455	PPSPNTILN-----TVFTVGSNKTSPK-----KCSYAVDGNNS	1488			
DB	1432	KLDNTIVLDWKNQTVGWHITICESFTNDVSIIINMGNI AALFLRBDPCVYLCSIA TD----	1487			
QY	1489	QGFOIFSSYQSSGMLDIDTGINNTDITKIVMAGSKTHT--FTASDHIA SLPANSFDMPT	1547			
DB	1488	-----IKIASSMIQ-----IQDKNISPLKXGSDILVELNAEDHVASKPSHESDPMVD	1537			
QY	1548	FKPLEIDASSLAFTNIIAPLDIVFETKAKDGRVLG-KIKQTL SVKRVNYPEDILFLRET	1606			
DB	1538	FNQVKVDLEGYDI-----PLVSEFI IKQPDGGYNDIVIESPIHIKLSKDTNSVLSLHKM	1592			
QY	1607	HSGAQYMOLGYYRIRLNTLLASQLVSRANTGIDTILTMETQRLPEPPLGEGFANFVLPK	1666			
DB	1593	PSGTQYMOIGYRTRKNTLFRKLAERANIGIDNVLSMETQNLPEPOLGEGFYAFKLP	1652			
QY	1667	YDPAHGBDERFKI HIGNVGGNGTQPYYSKMLSDTSETSMTLFVPYAEGYVHHGVRIG	1726			
DB	1653	YNKEHGBDERFKI HIGNIDGNSARQPYVEGMLSDI-ETT VTLFVPYAKGYI IREGVRIG	1711			
QY	1727	VGYKQITVNTWESAFPFDETKQFVLINDADHDSGMTQOGIVKNIKKYKGLNVSAT	1786			
DB	1712	VGYKKI IYDKWESAFPFYFDETKQFI FINDERADHDSGMTQOGIVKNIKKYKGLNVSAT	1771			

```

QY 1787 GYSAPMDFNSALYFWELFYTPMFCORLLOEKQFDEATQWVINYVNPAGYIYNGRIA 1846
D 1772 NNTBMDPENGANAIYFWELFYTPMFWFORLLOEQNFTESTRWLYIWNPAQSYSGEQW 1831
QY 1847 PWIWNCRPLEETTSNANPLDAIDDAVANDPMMYKATFWRLLDQLILRGDMAYRELT 1906
D 1832 DTYWNRVPLEEDTSNANPLDSVDPDAVAQHPMHYKVATFMKMLDLITRGDSAYRQLE 1891
QY 1907 RDLNEAKWYVRTLLELGDEPEDYSGQWAAAPSLSGAASQTVQAAQOOLTMLGRGVYS 1966
D 1892 RDTLNEAKWYVQALITLGDEPFYSLDNDWSEPRLEEAASQTRMHHYQHKMLQORAL 1951
QY 1967 KNLRTANSVLGLFLPEYNPALTDYQWTLRLRLFNRLNLSIDGQPLSLAIYAEPTDPKAL 2026
D 1952 PTKRTANSALTALFLQINKKLGQYQWTLTQRLYLNRLNLITDQPLSLSLVATPADPSML 2011
QY 2027 LTSWQASQGSVAFLPGTILSLYRPVPMLERTNVAQLTOGTSLLSMAEHDDADELTTL 2086
D 2012 LSAAITASQGGDLPHAVPMYRFPVILENAKGVSLIQFGNTLLSITERQDABALABI 2071
QY 2087 LLOQGMELATQIRIQORTVDVADIAVLAESRRAQNRLEKYQOQYDDEDINHGEORAM 2146
D 2072 LQOQSELALQSIKQDKMAEIDADKLALQESRGAQSRPDSFNTLIDEDVNAGEKQAM 2131
QY 2147 SILDAAAGQSLAGQVLSIAEGVADLVNPNVFLGACGSRGWAALRASAVMSLSATASQVS 2206
D 2132 DLYLSSSVLSTSGTALHMAAAADLVPMIYFPAVGSRFGALFNASAIETISASATRIA 2191
QY 2207 ADKISRSEAYRRRQEWIQRNADGEVQMDAQLESKIRREAAQMVQESTQOANTQ 2266
D 2192 ADKISQSEIYRRRQEWIQRNABEIKQIDAQATLAVRREAAVLQKYLETQOQATQ 2251
QY 2267 AQLELQKFTNKALYSWNRGKLSAIYQFPDLTQSFCLMAQEAULRRELTDNGVTFIRGG 2326
D 2252 AQLAFLQKFSNAALYNLWLRGLSIAIYQFDYDLAVSLCLMAEQTYQYELNNAAHFIKPG 2311
QY 2327 AWTGTTAGLMAGETLLALLAEMKVMLEDERDALEVRTRVSLAQFYQALSSDNFNLTEKL 2386
D 2312 AWHGTYAGLAGETLLMLAQMEKSYLEKDERALEVTRVSLAENVYAGLTENSFTLKDKV 2371
QY 2387 TOPLREGKGVNAGSVELKLSNRQTEASVRLSDIKIFSDYPESIGNTQLKQVSVTLPAL 2446
D 2372 TELVNAQSGSAGTTLNGLNVEGTOLQASLKLSLDLNIATDYPDGLGNTRIKQISVTLPAL 2431
QY 2447 VGPVEDIRAVLNYGGSIVMPRCSAIALSHGVNDGQFMDLDFNDKRYLPFEGISVNDGSGS 2506
D 2432 LGPYQDVRAILSYGGSTWMPRCCKAIALSHGVNDGSGQFQMDFNDAKLYLPFEGLPVADTGT 2491
QY 2507 LTLSPDADTRQKALLESLSLIIHLIRYTIYS 2538
D 2492 LTLSPFGISGKQKSLLSLIIHLIRYTIYS 2523

RESULT 7
AAV33729 standard; protein; 2522 AA.
XX
AC AAV33729;
XX
DT 09-NOV-1999 (first entry)
XX
DE DE
XX
XX Photorhabdus luminescens hph2-encoded insecticidal toxin.
XX
XX Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.
XX
XX Photorhabdus luminescens.
XX
XX WO942589-A2.
XX
PD 26-AUG-1999.
XX
PP 18-FEB-1999; 99WO-EP001015.

```

```

XX 20-FEB-1998; 98US-00027080.
PR 20-JAN-1999; 99US-0116439P.
XX
PA (NOVS ) NOVARTIS AG.
PA (NOVS ) NOVARTIS-ERFINDUNGEN VERW GES MBH.
PI Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MM,
Chen JS;
XX
XX MPI: 1999-527479/44.
DR N-PSDB; AAZ06831.
XX
XX New nucleic acid from Photorhabdus luminescens encoding insecticidal
PT toxins, used for making resistant transgenic plants.
XX
PS Claim 26; Page 133-140; 148pp; English.
XX
XX This sequence represents the hph2 gene encoded insecticidal toxin from
CC Photorhabdus luminescens. It is one of three insecticidal toxins
CC (AAV33728-Y33730) encoded by open reading frames (orfs) in a 38kb
CC fragment of P. luminescens DNA (AAZ06831). The hph2 gene was identified
CC using probe #2 which was amplified from P. luminescens genomic DNA using
CC PCR primers AAZ06829-Z06830. P. luminescens is a member of the
CC Enterobacteriaceae family and is a symbiotic bacterium of nematodes of
CC the genus Heterorhabditis. The nematodes colonise insect larvae, kill
CC them, and their offspring feed on the dead larvae. However, the
CC insecticidal agents are produced by P. luminescens rather than the
CC nematodes. The toxins have activity against Lepidopteran insects such as
CC Cabbage Looper (Trichoplusia ni), European Corn Borer (Ostrinia nubilalis)
CC and Fall Armyworm (Spodoptera frugiperda) and also against Coleopteran
CC insects (e.g., Colorado Potato Beetle, Leptinotarsa decimlineata). In
CC addition the toxins are active against strains resistant to known
CC insecticides. The DNA sequence can be used to generate transgenic plants
CC of various species that are resistant to economically important insect
CC pests and also for recombinant production of the toxins for use as
CC insecticides
XX
SQ Sequence 2522 AA;
XX
XX Query Match 39.5%; Score 5148.5; DB 2; Length 2522;
XX Best Local Similarity 43.0%; Pred. No. 0;
XX Matches 1121; Conservative 464; Mismatches 822; Indels 197; Gaps 55;
XX
QY 22 LADQLYSFLRKIFDDQLSWGEGARHLYHETIEQKNRNLLEARIFTRANPQLSGAIRL 81
D 28 LTDISHSSFNHFHQVSEHLSWEADLYHDAQQAQKDNRLYEARKLTNPQLQNAVHL 87
QY 82 GIERDSVSR-SYDEMFGARSSSVKPGSVASMFSPAGYLTLYREAKDLHFFSSAYHLDN 140
D 88 AIVAPNAELIGYNNQFSGRASQYVAPGTVSSMFSPAAYLTLYREARNLHASDSVYRLDT 147
QY 141 RRPDLADLTLSQSNMDEITLTLNELLLEHITRKTGDS-DALMESISTVYRQADITPY 199
D 148 RRPDLKSWALSQQNDMTSLSLSNELLLESIKTESKLDNTQVWMLSAFRPGATPY 207
QY 200 HQPYETIRQVIMTHDSTLSALSRNPEVMQAGASLLAILANISPELYNLTETEKNKA 259
D 208 HDAYENVKVIQLQDPGLEQLNASPAIAGLMHQASLLGINASISPELFNLTETEKNKA 267
QY 260 DALFAQNFSENIPTENFASQSWIAKYGLELSEVQKYLGMLOK-GYSDSTSAVYDNISGT 318
D 268 BELYKNFG-NIEPASLAMPPEYLRYYNLSDEELSQFICKASNFQOQESY--NQLITP 323
QY 319 LVVNNESKLEAYKIVTRVKTDDYDKNINFD--LMYEGNNQPFIRANFKVSREFGATLRKN 376
D 324 IVNSNDGTVKVIIRITR----EYTNANQVDVELFPYGGENTQVINKFQDSRDQDVSLSTK 379
QY 377 AGPSGIVGSLSGPLIANTNFKSNLNSISDSSEYKNGVKIYAYR-YTSSTSATNQGGIPT 435
D 380 LNDKRELIRIEGAPQVNIYESEHI--TLSTTDISQPFIEGLTRVYVPSSWA--YAAAKFT 435
QY 436 FESYPLTITFALKLNKAIIRLCITSGLSGSPNELOTVRSNDAQIINDSVLTQVFTLYFSHR 495

```

Db 436 IEEYNQVSLKLNKAIRLSRATLSPTILSISVRSVNOQLDINAELVGLVKFLTKYMQR 495
Qy 496 YALSFDDQVLNGSVINGYADDDSVSHFNRLFNTPPLKGIKEADGNTVSDIDPEBOSTF 555
Db 496 YAINAETALILCNALISQSYDNQPSQDFLNFNTPLNGQYFSTGDEIDLNGP-STGDM 554
Qy 556 ARSALMRGLGVNSGELYQLGKLAGVLDAONTITLSVFISSILYRLTLARVHOLTWNELC 615
Db 555 RKSVLKRAFNDIDTSLVRLKLTTHNNQDQGIKNLNLNLDLYIGKLAETHQUTIDELD 614
Qy 616 MLYGLSPNGKKTAS-LSSGELPLRVLWLYOVTOQLTEABITTEBAIMVLLCTPEFSGNIS 674
Db 615 LLL-VAVGEGETNLSAISDKQALALIKLNTITVWLQTKWSAQOLFVMTSTSYNKILTP 673
Qy 675 EISLNLNLRPSI---SEDMQAQSHNRELQABILAPFIAATLHLASPDMAWYILWNTNLR 731
Db 674 EIKNLLDVTYHGLQGFQDKDKAN-----LLHVMAPIYIAATLQLSSENVAHSVLLWADKJK 727
Qy 732 PGGDLIAGFWTL-VLKESINANET-----TQ---LYOFCHVMAQLSLSVQTLSEAE 780
Db 728 PGD-----GAMTAEXFMDWLNTQYTPDSSEVLATQEBHVOYCOALAEVMTYHSTGINEA 783
Qy 781 LSVLVISGFVILGAKNOPAGOHNDITLPSLYRFHOWINGLGNPGSDTILDLMLRQOQTADR 840
Db 784 FRLFWTKP-EMFGSSTEAVPAHDALSILMLTFADVWVALGEKASSVLAFAEANSILTAEQ 842
Qy 841 LASVGLDISMVTQAMVSA-----GVNQLQWODINTVLQWIDVASALHTMPSVI 890
Db 843 LADAMNLDANELLQAOSTQAOHQHLLPVVPTQKNFSCWTSIDTILQWVNAQLANVAPQGV 902
Qy 891 RTLNIYRVYALNKAESNLPSEWDEWOTLAEWNEAGLSQOQAOTLADYTAERLSSVLWNWF 950
Db 903 SALVGLDYI---QLNQKIPTYAQWESAGEILTAGLSQQADILHAFIDESRSAAALSTYY 958
Qy 951 LANTOPEGVLSHRDDLVSFLIDNOVSSAIKTRLAELAGIOLYINRALNRLEPNARA 1010
Db 959 IROVAKPAAALKSRDDLQYLLIDNOVSAALKTRIAELIASIOLYINRLEVENAHS 1018
Qy 1011 DVSTRQFPTDW-TVNNRYSTWGGVSRVLYPENYIDPTQRIQOTRMDELLENISQSCLS 1069
Db 1019 GVISKQFPIDWKYKRYSTWAGVSQLVYYPENYIDPTMRIGQTKWMDALLQSSQSOLN 1078
Qy 1070 RDTVEDAFKTYLTPETVADILKVYSAYHDNVSNLTGLTFWFGOTRENLPYVWRNVDSR 1129
Db 1079 ADTVEDAFMSYLTSEFQVANLKVISAYHDNINNOQGLTYFGLSETDTGEYIWRKSVDSK 1138
Qy 1130 MOAGELAAANAKWKTKIDTAVNPYKDAIRPVIFRERLHLIWEKEEVAK---NGTD--PV 1184
Db 1139 FSDGKFAANASEWHKIICPINPYRSTIRPVWYKSLYLLMLQKEITKQTSNGKQYQT 1198
Qy 1185 ETYDRFTLKLAFLRHGWSAPWSYDIIT---TQVEAVTDKKPDTERTLALASGFGEDTL 1241
Db 1199 ETDVRYELKLAHIRYDGTWNTPIITFDVNEKISKLEAKNKA-----GLYCAGYQGEDTL 1253
Qy 1242 LVFVYTKGSYSDPFGSNKNVAGMTIYDGGFGKMENTALSRYSOLKNTPIILHTQGNDL 1301
Db 1254 LVMFYNOQDTLDSY---KTASMOGLYIFADMEYKMDTGOYKSYRD--NSYKQFDTNS--- 1306
Qy 1302 VRKASYRAQDFEVPASLN--MGSALGDDSLTVMENGINIPOITTSKYSDDNLAI----- 1352
Db 1307 VRRVNNRYAEVYEIPSSVNSRKYDWDGYLYSMVYNGDIPITISYKATSSDLKIYISPKLR 1366
Qy 1353 TLHNAFTVRYDGSNGVIRNKQISAMKUTGVGDKSQYGNAPFI-----ANTVKH 1401
Db 1367 IHHNG-----YEQO-----QRNQCNLNNKYG-----KLGDKFIYVTSLGVNPNNSNKLWF 1412
Qy 1402 Y-----CGYSDLGCPITVYNTKNYIASVQGLHMAVYTRRLILTPVNNYIARLFEPF 1455
Db 1413 YPVQYQNGVNSGLSQGRLLFHRDNTYSSKVEAWIFGAGRS-----LT----- 1454
Qy 1456 FSPNTILNTVTVGS-NKTSDFPKCSYAVDGNNSQGFQIFSSYQSSGWLDDIDTGINNTDI 1514

Db 1455 -NPNAAGIDDVATDSLKNPNDLKQYVYMTD---SKG---TATDVSGPVDINTAISPAKV 1506
Qy 1515 KITWAGSKTHTFTASDHASLANSFPDAMPYTFKPLEIDASSLAFTNNIAPLDIVPETK 1574
Db 1507 QVTVKAGSKBOTFTADKNVSIQSPSPFDEMNYQNALEIDGSSLNFTNNSASIDITFTAF 1566
Qy 1575 AKDGRVLGKIQTLSVKRVNVPEDILFLRTHSGAQMQLGVYRIRLNTILLASOLVSA 1634
Db 1567 ABDGRKLGYSFSPITR-KVSTDNSLTLRHENGAQYMQMGVYRIRLNTILFARQLVARA 1625
Qy 1635 NTGIDTILMTBQRLPEPPLEGFFANPVLPKYDPABSHGDRWPKHIHGNVGGNTGROPY 1694
Db 1626 TTGIDTILSMETQIQSPOLCKGFYATFVIPPYNPSTHGDRWFKLYIKHVVDNNS-HII 1684
Qy 1695 YSGMLSDTSETSMTLFVP-----YAGGYMHGVLRGVGYQKI TYDNTWESAFYF 1745
Db 1685 YSGQLKDTN-ISTTLFPLDDVPINQDYSAKVM-----TFKKSPTSOTWGPPEVR 1735
Qy 1746 DETKQOQVLINDADHDSGMTQOQGIYKNIKKYKGLFNVSIATGYSAPMDFNSASALYYWEL 1805
Db 1736 DD--KGIVTN-----PKSILTHPESVNVNLNIS-----SEPMDFSGANSIYPWEL 1779
Qy 1806 FYTTPMCMFORLLQKOFDEATQMINVYNPAGYIVNGEIAPIWNCBPLBETTSWANP 1865
Db 1780 FYTTPMLVAQRLHEQNFDEANRWLKYVWSPGIVHGOIQNYQVNVNRPILLEDTSWNSDP 1839
Qy 1866 LDAIDPAVAQNDPMHYKIATFMBLLDOLILRGDMAYRELTRDALNEAKMAYVRLLELG 1925
Db 1840 LDSVDPVAQAQNDPMHYKSTFMTDLIIARGDHAYQOLERDITLINEAKMYMQALHULG 1899
Qy 1926 DEPEDYSGQOWAAPSLSGAASQTVQAAVQOQLTMLGRGVGS-KNLRTANSILVGLFPEYN 1984
Db 1900 DKPYLPSTTWNDPRLDKAADITQSAHSSISVALRQSTPALLSLRSANTLTLDLFLPOIN 1959
Qy 1985 PALTDYQOTLRLRFLNLRHNLSDGQPLSLAIYAEPTDPKALLTSMVQASOGGSAILPBT 2044
Db 1960 EVMANYQTLAQRVYVNLRHNLSDGQPLYLPIYATPADPKALLSAAVATSGGGKLPBSF 2019
Qy 2045 LSLYRFPVMLERTNLVNAQLTOFGTSLLSMAEHDDADELTLLLLQOQGMELATQSIROQR 2104
Db 2020 MSLRFPMLHNAKMSVSQLTQFGSTLQNIIRERQDAEALNALLQNAELILTNLSIODK 2079
Qy 2105 TVDEVDADIAVLAESRSQAQNRLEKYQOLYDEDINHGEQRAMSLDLDAAGOSLAGQVLSI 2164
Db 2080 TIEELDAEKTVLEKSKAGAQSFDSYSKLHDENINAGENQAWTLRASAAGLTTAQASRL 2139
Qy 2165 AEGVADLVPNVFGIACGGSRWGAALRASVMSLSATASQYSADKISRSEAYRRRRQWE 2224
Db 2140 AGAAADLVPNIFGPGAGGSRWGAIAEATGYVMEFSANVMNTEADKISQSEYRRRRQWE 2199
Qy 2225 IORNDAGEVKKOMDAOLESKIRREAAOMOVEYOETQOAHQAOQLLELQKFTNKALYSW 2284
Db 2200 IORNAEABLQOLDAQLKSLAVRREAVALQTSUKTQOEQOQAQLAFLQRFPSNOALTNW 2259
Qy 2285 MRGKLSAIYQYFQDLTOSFCLMAQALRRELTDNGVTEIRGGAWNGTTAGLMAGETLLN 2344
Db 2260 LRGLAALYFQYDLATARCLMAQAYWEISDDSAFIPKPGAMQGTYAGLAGETLMS 2319
Qy 2345 LAEMEKWLERDERALEVTRTVSLAQFYQALSSON--FNLTETKLTQFLREKGNVGASGN 2402
Db 2320 LAQMEDAHLRRDKALEVERTVSLAEIYAGLPQDKGPFSLAQEIEKLVNAGSGSAGSGN 2379
Qy 2403 ELKL-----SNRQIEASVRLSLDKITPSDYPSELGNTRQLKOVSVTLPALVGPYEDIRAVL 2457
Db 2380 NLAFAGAGTDTKTSLOASISLADLKIREDPESIGKIRIKQLSVTLPALGPGYQVQAIL 2439
Qy 2458 NYGSGIVNPRGCSAIALSHGVNDSQGFMLDENDSRFLPEFEGISVNDSGSLTSPFDAT-- 2515
Db 2440 SYGDKAGLAGCAALAVSHGINDSQGFOLDENDCKFLFPEGIAI-DOGTTLTSPFNASTP 2498
Qy 2516 --DRQKALLESIDILHIRYTIIR 2537
Db 2499 AKGQOATMLKTLNDIILHIRYTIK 2522

RESULT 8

AAW56572
ID AAW56572 standard; protein; 2516 AA.

XX AC AAW56572;

DT 07-AUG-1998 (first entry)

XX Toxin TcdA, encoded by the tcdA gene from genomic region tcd.

XX Photorhabdus luminescens W-14; nematode; symbiotic; Heterorhabdus; tca;
KW tcd; tcd; insecticidal activity; toxin; Lepidoptera; Coleoptera;
KW Hymenoptera; Diptera; Dictyoptera; Acarina; Homoptera; Southern;
KW Western corn rootworm; Colorado potato beetle; mealworm; boll weevil;
KW turf grub; beetle armyworm; black cutworm; cabbage looper; codling moth;
KW corn earworm; European corn borer; Tobacco hornworm; budworm.

XX Photorhabdus luminescens.

XX W09808932-A1.

XX 05-MAR-1998.

XX 05-MAY-1997; 97WO-US007657.

XX 28-AUG-1996; 96US-00705484.

XX 06-NOV-1996; 96US-00743699.

XX 06-NOV-1996; 96WO-US018003.

XX (DWC) DOWELANCO.

XX (WISC) WISCONSIN ALUMNI RES FOUND.

XX Ensign JC, Bowen DJ, Petell J, Fatig R, Schoonover S;

XX Ffrench-Constant RH, Rocheteau TA, Blackburn MB, Hey TD, Merlo DJ;

XX Orr GL, Roberts JL, Strickland JA, Guo L, Ciche TA, Sukhapinda K;

XX WPI; 1998-179427/16.

XX N-PSDB; AAV29928.

XX Isolated toxins from Photorhabdus luminescens strains - useful for

XX control of insect pests.

XX Claim 34; Page 224-231; 321pp; English.

XX The present sequence represents a protein named TcdA of the bacterium
XX Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the
XX nematodes of the Heterorhabdus genus. The bacterium has at least 4
XX distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are
XX produced from these regions that are associated with insecticidal
XX activity. The native toxins are secreted proteins. The proteins are toxic
XX to insects upon exposure and especially when ingested. The nucleic acid
XX sequence can be used to produce transgenic plants, baculoviruses or
XX microbial hosts for toxin production. They can be used to control insects
XX pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera,
XX Dictyoptera, Acarina or Homoptera orders, especially the Southern or
XX Western corn rootworm, Colorado potato beetle, mealworm, boll weevil,
XX turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth,
XX corn earworm, European corn borer or tobacco hornworm or budworm

XX Sequence 2516 AA;

XX Query Match 38.7%; Score 5043; DB 2; Length 2516;

XX Best Local Similarity 42.6%; Pred. No. 0;

XX Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;

XX 22 LADLOVSESELRKIFDDQLSWGEARHLYHETIEQKNNRLEARIFRANPQLSGAURL 81

XX 21 LTDISHSFNEFRQVSEHLSWSETHLDYHDAQQAQKNNRLEARILKEANPQLQNAVHL 80

XX 82 GIERDSVSR-SYDEMFGARSSFKVPGSVASMPSPAGYITELYREAKDLHFSSSAYHLDN 140

Db 81 AILAPNAELIGYNNQFSGRASQYVAPGTVSSMPSPAAYTELYREARNLHASDSVYLD 140

Qy 141 RRPDLADLTLSQSNMDEISTLTLSNELLEHI-TRKTCGSDSALMESISTYRQAIDTPY 199

Db 141 RRPDLKSMALSQNNMDELSTLSNELLESIKTESKLENTKYMELMSTRPSPATPY 200

Qy 200 HOPYETIRQVIMTHDSTLSALSRNPEVMQAGASLLAILANISPELYNILEETEEKNA 259

Db 201 HDAYENVREVIQLODPGLEQLNASPAIAGLMHQASILLGINASISPELFNILEETEGNA 260

Qy 260 DALFAQNFSENITPENFASQSWIAKYGYGLESEVOKYLGMLON-GYSDSTSSAYVDNISTG 318

Db 261 ELYIKKFG-NIEPASLAMPYELKRYNLSDELQFQKASNFQOQESVSN---NQLITP 316

Qy 319 LYVNESKLEAYKTR-VKTDYDKNINYFDLMYEGNNOFFIRANPKVSRFECAT----- 372

Db 317 VNSSDGTGVKVRITREYTNAYQMDVELFP--FGEN---YRLDYKFNFNYSYLSIK 371

Qy 373 -----LRKNAGPSGIVGSLGPIANTNFKSNVLSNLSDEYKNGVKIYAIRYTSSTS 425

Db 372 LNDKRELVRTEGAQ-----VNIYSANITLTAD--ISQPFEGITRVLPSGS 418

Qy 426 ATNCGGIFTPESSYPLTIFALKKAIKRLCLTSLSPNELOTIVRSDNAQGIINDSVLTK 485

Db 419 WA-YAAAKFTVEYNQYSFLLKNAIKRLSRATELSPITLEGIVRSVNLQDINTVLGK 477

Qy 486 VFYTLFYSHRYALSFDQAQVLSNGSVINQVADDSVSHFNRLFNTPPLKGIKEADGNTVS 545

Db 478 VFLTKYVMQRYAHAEATALLCNAPISQSYDNQPSQDFDLFNTPLLNQYFSTGDEED 537

Qy 546 IDPDEQSTFARSALMRGLGVNSGELYQLGKAGVLDQAQNTITLSVFVLSLYRLTLAR 605

Db 538 LN-SGSTGDRKTIKRAFNDVDSFLRLKITDHNKDGKIKNNLKNLSNYIGKLLAD 596

Qy 606 VHLTVNELCMYGLSPFNGKTTAS-LSSGELPRVIMLYQVOTMLTEAITTEALWLC 664

Db 597 IHLTIDELDL-FAVGEKTNLSAIDSKQATIRKLTITSLWHTQKSVQFQFIMT 655

Qy 665 TPEFSGNISPEISNLNLRPSI---SEDMAOSHRELQAEILAPPIAATLHLASPDMA 721

Db 656 STSYNKTITPEIKNLLDTVYHGLQGFQDKAD-----LLHYMAPYIAATLQSSENV 709

Qy 722 YILLWTDNLRPGGLDIAGFMT-----LVLESINANETTO-----LVQFCHVMAQLS 770

Db 710 SVLLWADKLQGD-----GAMTAEFKFDWMLNTKYTFEGSSEAVETQEHIVQYCALA 765

Qy 771 VOTLRLSRAELSVLVSIGFAVLGAKNQAPAGOHNIDTLPFLYRHHOWINGLNGPSTLDM 830

Db 766 YHSTGINENAFRLFVTKP-EMFGAATGAAPAHADALSIMLTRPADWVNALGEKASVLA 824

Qy 831 LRQQTLTADRLASVMGLDISMVTQAMVSA-----GVNOLQCMQDINTVLOWIDVA 880

Db 825 FEANSLTAEQLADANLNDANLLQASIQAQNHQHLPPVTPPENAFSCWTSINTILQVWVA 884

Qy 881 SALHTMPSVIRTVNIRVITALKAESNLPSHDEWQTLAENWEAGLSTQAOATLADYTA 940

Db 885 QQLNAPQGVSAVGLDYIQSMKET---PTYAQWENAGVLTAGLNSQANTLHAFLE 940

Qy 941 RLSSVLCHWFLANIQEGVSLHSDRLYSYFLIDNQVSSAIKTTTLABAIGQLVYINRA 1000

Db 941 SRSAALSTYIRQVAKAAAAIKSRDDLYOYLLIDNQVSAIKTTIRIAEIASIQLVYINRA 1000

Qy 1001 LNRIEPPNARADVSTRQFFTDW--TVNNRYSTMGVSRVLYPENYIDPTORIGQTRMDEL 1059

Db 1001 LENVEENANSVYISQFFIDWDKYNKRYSTWAGVSQLVYYPENYIDPTMTRIGQTKMDAL 1060

Qy 1060 LENISQSKLSDRTVEDAFKTYLTRTETVADLVKVSAYHDVNSNTGLTWFGQTRNLPE 1119

Db 1061 LQSVSQSLNADTVEDAFMSYLTSPQVANLKVISAYHDNINNDQGLTYFGLSETDAGE 1120

Qy 1120 YWVRNVDYSRMCAELAANAKWKTKIDTAVNPKYDAIRPVIFFRELHLIWEKEVAK- 1178

Db 1121 YWVRSDHSHKFDGKFAANAWSEWHKIDCPINPYKSTIRPVYKSLYLLWLEQKEITQ 1180

QY 1179 --NGTD--PVETVDRFTLKLAFRLHDSGSAPWSYDITTOVEAVTDKPKDTERLALAASG 1234
 DB 1181 TGNKSGQYOTEDRYELKJAHIRYDGTWNTPTFDVNNKISL--KLEKNRAPGLYCAG 1238
 QY 1235 FQGEDTLVAVYKTKSGSDFGSGMKNVAGMTIYDGSFKKMENTALSRYSQLKNTFDII 1294
 DB 1239 YQGEDTLVAVYKTKSGSDFGSGMKNVAGMTIYDGSFKKMENTALSRYSQLKNTFDII 1294
 QY 1295 HTQGNDLVRKASYRPAQDFEVPASLNMG--AIGDSDLTVMENGINPOITSKYSDNLAI 1352
 DB 1295 DTNN--VRRVNNRYAEDEIPESSVSRKDYGLWSVYNGDIPNTINKAASSDLKI 1351
 QY 1353 -----TLHNAFTVRVDSGNVIRNKOISAMKLTGVDSKSOYCNARFIANTVKHYGGY 1405
 DB 1352 YISPKLRIHNG-----YEGQ-----KRQCNLMNKYG-----KLGDKEFIV-----Y 1388
 QY 1406 SDLG-GRITVYVTKRNY--IASVOGHLMNADYTRRLILTPVENNYARLFEF--PFSNPTIL 1462
 DB 1389 TSLGWNPNSSKLMFYVYQYSGNTSLNQGR--LLPHRTTYTPSKYEAWIPGAKRSLT 1446
 QY 1463 NTVFTVGS-----NKTSDFKKCSYAVDGNNSQGFQIFSSYQSGWLDIDTGINNTDIK 1515
 DB 1447 NQNAAGDDYATDSLKNKDLKQYIFMTD---SKG---TATDVSGPVEINTAISPAKVQ 1499
 QY 1516 ITVMAGSKTHFTASDHASLPAISFDPAMPYTFKPLEIDASSLAFTNRIAPLDIVFETKA 1575
 DB 1500 IIVKAGGKEQFTADKQVSIQPSFSDENMYQFNALIEDGSLGNFINNSAIDVTFTAPA 1559
 QY 1576 KQGRVLG---KIKOTLSKRVNVPEDILFLRETHSGAQYMQLVGYRIRLNTLLASQV 1631
 DB 1560 EDGRKLGYESPISIVTLKV-----STDNALTLLHNEGAQYMQVQSYRTRLTNLFARQV 1614
 QY 1632 SRANTGIDTILMETQRLEPPPLGSGFPANFVLPKYDPAEHGDERWFKIHXGNVGNCTGR 1691
 DB 1615 ARATGIDTILSMETQNIQEPQLGKGFYATFVIPPYNLSHGDREWFKLYKHVVVDNNS- 1673
 QY 1692 QPYISGMLSDTSETNTLF-----VPYAEGYMHGVRGLGVGYKITYDNTWESAFFFD 1746
 DB 1674 HIIYSQGLTDTN-INITLFIPLDDVPLNQDHY-----AKYMTTFKPSDGTWGWPHFVRD 1728
 QY 1747 ETQKQFVLINDADHDSGMTQOQIVNKKYKGLFNLVSIATGYSAFMDFNSASALYWFELF 1806
 DB 1729 D--KGIVIN-----PKSILTHFESVNVNINIS-----SEPMDFSGANSLYFWELF 1772
 QY 1807 YPTPMCFQRLQEQFDEATQWIVYVNPAGYVINGEIAPIWNCRPLEETTSNANPL 1866
 DB 1773 YPTPMLVAQRLLEHQNFDEANRWLYVWSPGYIVHGQIQYQWNVVRPLEEDTSWNSDPL 1832
 QY 1867 DAIDPDVAQNDPMHYKIATFNRLLDQILRGDMAYRELTDRALNEAKWYVTRILELGD 1926
 DB 1833 DSVDPDAVAQNDPMHYKYSTFNRLLDQILRGDMAYRELTDRALNEAKWYVTRILELGD 1892
 QY 1927 EPEDYGSQQAAPSLSGAASQVQAAQODLTMKRG--GVSXKLRTANSILVGLFLPEY 1983
 DB 1893 KPYLPLSTWSDPRLDRAADITQNAHSAIVALAQNIPTPAPLSIRSANLTDLFLPQI 1952
 QY 1984 NPALTDYQOTLRRLRLEHNLSDIQOPLSLAIYAEPDTPKALLTSMVOAGSGGSAVLPG 2043
 DB 1953 NEVMNMYQTLAQRVYVNLRLHNLSDIQOPLSLAIYAEPDTPKALLTSMVOAGSGGSAVLPG 2012
 QY 2044 TSLYRFPVWMLERTRNLVALQTOFTGTSLSMAEHDDADELTLLLOQGHETATQIRIQO 2103
 DB 2013 FMSLWRFPMLHNAEGRWYSQLTFQSTGLQNIITIERQDAEALNALLQNAEALTLTSLIQD 2072
 QY 2104 RTVDSDVDADIAVLESRSQAQRLEKYQOLYDIEDINHGEORAMSLIDAAAGOSLAGQVLS 2163
 DB 2073 KTIIELEDAEKTIVLEKSKAGQQRFPYSYKGLYDENINAGENQANTLRSAGLITTAQASR 2132
 QY 2164 IAEGVADLVPNVFLACCGSRWGAALRASASVMSISATASQVSAQYADKISRSEAYRRRROR 2223
 DB 2133 LAGAAADLVPNIFGPGAGGSRWGAATAEATGYVMEFSANVMNTAEADKISQSEYRRRROR 2192

QY 2224 EIORDNADGEVKQMDAQLESKIRREAAQMOVYEQETOQAHTOQAELELORFKFTNKALYS 2283
 DB 2193 EIORNNAEALKQIDAQLKSLAVREAAVLQKTSIKTQEQEQSOLAFLOKFKFSNALYN 2252
 QY 2284 WMRGKLSAIYQFPDLTQSFCLMAQEAALRRLTONGVTPIRGGAWNGTTAGLMAGETILL 2343
 DB 2253 WLRGRLAAIYQFPYDLAVARCLMAEQAYRWELNDSARFIKPGAWQGYAGLLAGETJML 2312
 QY 2344 NLAWEKWLRLERDERALEVTRTVSLAQFYQALSSDN--FNLTEKLTQFLREGKGNVGASG 2401
 DB 2313 SLAQMEDAHLKRDKRALEVERTVSLAEVYAGLPKDNQGFSLAQEIDKLVSGSGSAGSGN 2372
 QY 2402 NELKL-----SNRQIEASVRLSDLKIFSDYDESIGNTROLKQVSVTLPALVGPYEDIRAV 2456
 DB 2373 NNLAFGAGTDTKTSLOASVSFADLKIREDPASLGKIRRIKQISVTLFALLGPYQDVQAI 2432
 QY 2457 LNYGGSIVMPGCGSAIALSHGVNDSGQFMDLDFNSRYLPPFGISVNDSSGSLTSLSPDAT- 2515
 DB 2433 LSYGDKAGLANGCEALAVSHGVNDSGQFMDLDFNSRYLPPFGISVNDSSGSLTSLSPDAT- 2515
 QY 2516 ---DROKALLSLSLIIILHRYTIR 2537
 DB 2492 PEKQOATMLKTLNDIILHRYTIR 2516
 RESULT 9
 AAB72609 standard; protein; 2516 AA.
 AC AAB72609;
 DT 04-MAY-2001 (first entry)
 DE Photorhabdus tcda toxin.
 KW Tcda; Tcda; insect toxin; plant; insect resistance.
 OS Photorhabdus sp.
 FN WO200111029-A1.
 XX 15-FEB-2001.
 PF 11-AUG-2000; 2000WO-US022237.
 PR 11-AUG-1999; 99US-0148356P.
 PA (DWC) DOW AGROSCIENCES LLC.
 PI Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW;
 PI Sukhapinda K, Merlo AO;
 XX WPI; 2001-191536/19.
 DR N-PSDB; AAF58778.
 PT Novel polynucleotide sequence encoding insect toxins, useful for
 PT producing transgenic plants having resistance to insects, especially corn
 PT rootworm.
 PS Disclosure; Page 51-61; 106pp; English.
 CC The present invention provides the protein and coding sequences of
 CC modified versions of the photorhabdus Tcda and Tcda toxins. These are
 CC suitable for expression in plants. The toxins are effective against
 CC insects upon ingestion, and the sequences provided can be used to produce
 CC transgenic plants with insect resistance. The present sequence is the
 CC photorhabdus Tcda protein
 XX Sequence 2516 AA;
 SQ
 Query Match 38.7%; Score 5043; DB 4; Length 2516;
 Best Local Similarity 42.6%; Pred. No. 0;
 Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;

Db 766 YHSTGINENAFRLFVTKP-EMFGAATGAAPAHADALSILMLTRPADVWVALGKASVSLAA 824
 QY 831 LRQOTLTADRLASVGLDLSMTVQAMVSA-----GVNQLQCWDINTVLOWIDVA 880
 Db 825 PEANSITAEQLADAMNLDANILLOQASIQAHQHLPVTPPENAFSCWTSINTILOQWVA 884
 QY 881 SALHTWPSVIRTLVNIYVYALNKAESNIPSWDEWQTLAENMEAGLSTQAOATLADYTA 940
 Db 885 QQLNVAPOGVSAVGLDVIQSMKET-----PTVAQMENAAAGVLTAGLNSQOANTLHAFLE 940
 QY 941 RLSSVLCWFLANIQEFGVLSHSDLDLYSVFLIDNQVSSAIKTRLAABAIAGIOLYNRA 1000
 Db 941 SRSAALSTYYIROVAKAAAIKSRDLYQYLLIDNQVSSAIKTRIAEAIASIQLYNRA 1000
 QY 1001 LNRIBENARADVSTROFFTDW-TVNNRYSTWGVSRVLYVPENYIDPTORIGOTRMWDEL 1059
 Db 1001 LENVENANSGVTSROFFTDWYKRYSTWAGVSQVLYVPENYIDPTWRIGOTRMWDEL 1060
 QY 1060 LENISQSLSRDVEDAFKYLTRPETVADLVKVSAYHDVNSNTGLTFWVGQTRLENPE 1119
 Db 1061 LOSVSQSLNADTVEDAFMSYLTSPQVANLKVISAHDNINNDQGLTFVIGLSETDAGE 1120
 QY 1120 YWRNVYDISRMQAGELAAWAKWTKIDTAVNPYKDAIRPVIPIREHLHIWKEBEVAK- 1178
 Db 1121 YWRSYVSHSKFNDGKFAANAWSEWHKIDCPINPYKSTIRPVIYKSLYLLWLEQKEITKQ 1180
 QY 1179 --NGTD--PVETVDRFTLKLAFURHDSWAPSMDYITQVEAVTDKPDTERLALASG 1234
 Db 1181 TGNKDCGYOTEDYRELKLHRIYDGTWPTPIFDVNVKISEL--KLEKNRAGLYCAG 1238
 QY 1235 FQGEDTLVYVYKTKSGSDFGSNKNVAGMTIYDGSFKKMENTALSRYSLKNTFDII 1294
 Db 1239 YQGEDTLVWYFVNOQDLSY--KNASMOGLYIFADMAKMDTPEQSNVYRD--NSYQOF 1294
 QY 1295 HTQGHDLVRKASVRFPAQDEVPASLNMGS--ALGDLSLTMENGNIPQITSKYSSDNLA 1352
 Db 1295 DTNN---VRRVNNRYAEDEIPESSVSSRDKYDGMGYLLSMVYNGDIFTINYKAASDLKI 1351
 QY 1353 -----TLHNAFTVRYDGSNGVIRNKQISAMKLTGVGDKGYGNAPFIANTVRYGY 1405
 Db 1352 YISPKLRIHNG-----YEQ-----KRNQCNLMNKG-----KLGDKFIV-----Y 1388
 QY 1406 SDLG-GPITVYNTKRY--IASVOGHLMNADYTRRLILTPVNNYIARLEP-PPFNTIL 1462
 Db 1389 TSLGVNPNNSNKLMPYVYQYSGNTSGLNQGR--LLFHRDTTPSKVEAWIPGAKRSLT 1446
 QY 1463 NTVFTVGS-----NKTSDPKCSYAVDGNNSQGFQIFSSYQSSGWLDDIDGINNTDIK 1515
 Db 1447 NQNAAGDDYATDSLKNKPDLLKQYIFMTD---SKG-----TATDVSGPVEINTAISPARKV 1499
 QY 1516 ITVMAGSKTHFTASDHASLPANSPDAMPYTFKPLEIDASSLAFNTNIAPLDIPVETKA 1575
 Db 1500 IIVKAGGKEQTTFADKXDSIQSPSPDEMNYQFNALEIDGSLNFNNNSASIDVTFAPA 1559
 QY 1576 KQGRVLG-----KIKOTLSYKRVNYPEDILFRETSHGSAQYMQLGVIYRIRLNTLASQLV 1631
 Db 1560 EDGRKLYGSEFSIPVTLKV-----STDNALTLHNEGAQYMQWQSYRTRLNTLFAQLV 1614
 QY 1632 SRANTGIDILMTETQRLPEPLPGSFPANFVLPKYDPAHGDGRWFKHIGNVGNTGR 1691
 Db 1615 ABATTGIDILMETQNIQEPQLGKGFYATFVIPPYNLSTHGDGRWFKLYIKHVVDNNS- 1673
 QY 1692 QPYVSMGLSDTSETSMTLR-----VPYAEYGYMHGVLGVQKITYDNTWESAFFYD 1746
 Db 1674 HIIYSQQLTDTN-INITLFIPLDDVPLNQDYH-----AKYMTFKKSPSGTGWGPHFVTD 1728
 QY 1747 ETQKQVLLINDADHDSGMTQOGIVKNIKKYKGLFNVSATGYASAPMDFNSALYWFEL 1806
 Db 1729 D--KGIVTIN-----PKSILTFHESVNVNINIS-----SEPMDFGANSLYWFEL 1772
 QY 1807 YTPMCMCFORLLOEQFDEATQWINTVNPAGYINGETAPWIWNCRPLEETTSNANPL 1866

Db 1773 YTPMLVLAQRLLHEONFDEANRWLKYVSPSPSYIVHGQIQNTQWNVRLPLEDTSWNSDEL 1832
 QY 1867 DAIDPDVAQNDDPMHYKIATFNRLDQILRIGDMAYRELTRDALNEAKWYVTRTELLGD 1926
 Db 1833 DSDVDPDAVAQHPDMHYKYVSTFWRTLDLLIARGDHAYRQLERDTLNEAKWYVQALHLLGD 1892
 QY 1927 EPEDVGSQWAAAPSLSGAASQTVQAAYQODLTWMLGRG---GVSKMLRTANSVLGLPLPEY 1983
 Db 1893 KEYPLSTTWSDFRLDRAADITTONAHSASIAVALRQNIPTPAPLSLRSAANTLTDLFLPOI 1952
 QY 1984 NPAALDYWOTLRLRLFNLRHNLSDIOPLSLAIYAEPDTPKALLTSMVQASOGGSVAVLPG 2043
 Db 1953 NEVMWNYWOTLAQRVYNLRHNLSDIOQPLYLPIYATPADPKALLSAAVATSOGGKLPS 2012
 QY 2044 TLSLYRFPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTLLTLLQOQGMELATQSIROO 2103
 Db 2013 FNSLWRFPHMLENARGWVSQLTQFGSTLQNIITERQDAEALNALLQNAEELILTLSIQD 2072
 QY 2104 RTVDEVDADIIVLAESRRSAQNRLEKYQOQYDIEDINHGEORAMSLDDAAAGOSLAGOVLS 2163
 Db 2073 KTEBELDAEKTVLEKSKAGAQSRFDSYGLKYDENINAGENQAMTLRASAAGLTAVQASR 2132
 QY 2164 IAEGVADLVPNVFGACGSRWGAALRASAVMSLSATASOYSADKISSEAYRRRRQSW 2223
 Db 2133 LAGAAADLVPNIFGAGGSRWGAIAEATGYVMEFSANVMTEADKISQSEYRRRRQSW 2192
 QY 2224 ETQRONADCEVKOMDAQLESKIRREAAQMVQEOYQTOQAHTQAQLELLQKFTNNKALYS 2283
 Db 2193 ETQRNNAEALKQIDAQLKSLAVREAAVLQKTSKTQEQTSQSLAFLOKFSNOALYN 2252
 QY 2284 WNRKLSAIYQFFDLTOSFCMLQAERRELTNGVTPIROGAWNTTLAGLMAGETLLL 2343
 Db 2253 WLRGLAAIYFOYDLAVARCLMAEQAVRWELNDSARFIKPGAWQGTYAGLLAGETML 2312
 QY 2344 NLAEMEKWLERDERALEVTRTVSLAOFYQALSSDN--FNLTEKLTQFLREGKGNVGASG 2401
 Db 2313 SLAQWEDAHKRRKALEVERTVSLAEVAGLPKONGPSPSLAQIBDKLVSSQSSGASGN 2372
 QY 2402 NELKL-----SNRQTEASVRLSDLKIFSDYPESLGNTRQLQKQSVTLPALVGPYEDIRAV 2456
 Db 2373 NNLPAGAGTDTKTSLQASVSPADLKIRREDYPASLKKIRRIKQISVTLPALGYPQDVQAI 2432
 QY 2457 LNYGSIWMPRGCSAIALSHGVNDSQWMLDNDSRYLPFEGISVNDSSGLTSLSPDAT- 2515
 Db 2433 LSYGDKAGLANGCEALAVSHGVNDSQWMLDNDSRYLPFEGISVNDSSGLTSLSPDAT- 2491
 QY 2516 ---DROKALLSLSDIILHRYTIR 2537
 Db 2492 PBKQKQWMLKTDLNDIILHIRTIK 2516

RESULT 11

ADP18614

ID ADP18614 standard; protein; 2516 AA.

XX AC ADP18614;

XX XX ADP18614;

DT 12-AUG-2004 (first entry)

XX DE Photorhabdus luminescens W-14 tcdal protein SEQ ID NO:2.

XX XX tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;

KW orally active insect toxin; insect toxin; tcdal.

XX OS Photorhabdus luminescens.

XX XX WO200404217-A2.

PD 27-MAY-2004.

XX XX 12-NOV-2003; 2003WO-1B005553.

PF 12-NOV-2002; 2002US-0425672P.

XX XX

XX (UYBA-) UNIV BATH.
 XX Ffrench-Constant RH, Waterfield NR;
 XX WPI: 2004-411735/38.
 XX N-PSDB; ADP18613.
 XX New isolated Photorhabdus luminescens nucleic acids, useful for
 PT expressing orally active insect toxin or for generating transgenic plants
 PT with enhanced resistance to insects.
 XX Disclosure; SEQ ID NO 2; 118pp; English.
 XX The present invention describes DNA sequences from the tcd genomic region
 CC of Photorhabdus luminescens W-14. Also described: (1) a transgenic
 CC monocot or dicot cell having a genome comprising a nucleic acid sequence
 CC that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant
 CC with a genome comprising a nucleic acid sequence that encodes SEQ ID
 CC NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a
 CC progeny of seed; (5) a method of producing toxin A of Photorhabdus
 CC luminescens W-14 in a heterologous host; and (6) a method of producing an
 CC orally active insect toxin. The nucleotide sequences are useful for
 CC heterologous expression of orally active insect toxin. They can also be
 CC used for generating transgenic plants with enhanced resistance to
 CC insects. The present sequence represents Photorhabdus luminescens W-14
 CC tcdA1, which is given in the exemplification of the present invention.
 XX Sequence 2516 AA;
 Query Match 38.7%; Score 5043; DB 8; Length 2516;
 Best Local Similarity 42.6%; Pred. No. 0;
 Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;
 22 LADLOVLFSELRKIFDDQLSGEAREHLYHETIEQKNNRLEARIFRANPQLSGAIRL 81
 21 LTDISHSEFNEFRQVSEHLSWSETHLDYHDAQQAQKQNRLEYEARILKRAPQLQNAVHL 80
 82 GIERDSVR-SYDENMFGARSSFFVPGSVASMPSPAGYLTLYERAKDLHFSSSAYHLDN 140
 81 AILAPNAELIGYNNQFSGRASQYVAPGTVSSMFPSPAAVLTLYERARNLHASDSVYLD 140
 141 RPDLDLTLSONMDTETSTLTLSNELLLEHI-TRKTGSDALMESLSTVROAIDTPY 199
 141 RPDLSKWSALSOQNDIEUSTLSNELLLESIKTESKLENTKTVKWEMLSTFRPGATPY 200
 200 HOPYETIROVIMTHDSTLSALSRNPEVMQABGASLAILANISPELYNLTETEITKNA 259
 201 HDAYENVREVIQDPGLEQLNASPAIAGLMHQASLILGINASISPELFNLTETEITGNA 260
 260 DALFQNFSENITPENFASQSWIAKYGLESEVOKYGLMLQN-GYSDSTSAVDNISTG 318
 261 EELYKKNFG-NTEPASLAMPYELKRYNLSDELSQFTIGKSNFQOQYSN---NQLITP 316
 319 LVVNNEKLEAVKTR-VKTDYDKNINFDLMYGNNOFFFRANFKVSRFEGAT- 372
 317 VNSSDGVKVRITREYTTNAYQMDVELFP--FGEN---YRLDYKKNFNASVLSIK 371
 373 -----LRKNAGPSGIVGSLSPLIANTNFKSNLSNISDSEYKNGVKIYAYRTSSTS 425
 372 LNDKRELVRTEGAPO-----VNIEYSANITLTAD--ISQPFELGLTRVLPSSG 418
 426 ATNQGGLITFTFESYPLTIFALKKNAIRCLTSGLSPELQTVIRSDNAQGLINDSVLTK 485
 419 WA-YAAAKFTVEYNQYSFSLKNAIRLSRATSELSPTILEGIVRSVNLQDINDTVLKG 477
 486 FVYTLFYSRYALSDPDAOVLSGVNOVADDDSVHFNRLNFTPLKGIPEADGNTVS 545
 478 VFETKYMQRYAHAEITALILCNAPISQBSYNDQSPQDFRLNFTPLANGQYESTGDEID 537
 546 IDPBEQSTFARSALMRGLGVNSGELYQGLAGVLDAQNTITLSVFVSSLYRUTLLAR 605
 538 LN-SGSTGDRKTLKRAFNIDVSLFRLLKTTDHNKDGKIKNNLKNLSNLYIGKLLAD 596

606 VHQLTVNELCWLGLSPNGKTTAS-LSSGELPRLVILWLYOVTOWLTEAITTEAILWLLC 664
 597 IHQUTIDELDLLL-IAVGEGKTNLSAISDKOLATLIRKLTNTITTSWLHTQKWSVQLFTMT 655
 665 TPEFSGNISPEISNLLNNLRPSI---SEDMAQSHNRELQABEILAPFIAATLHLASPDMA 721
 656 STSYNKTLTPEIKNLLDTVYHGLQFPDKAD-----LLHVMAPYIAATLQLSSENAH 709
 722 YILWNTNLRPGGLDIDAGFWT-----LVUKESLNANETQ-----LVQFCHVMAQLSLS 770
 710 SVLLWADKLPQGD---GAMTAERKFWDLNLTYPGSGSEAVETQEHIVQYCOALAQLSMV 765
 771 VQTLRLSEAELSVLVSGFAVLGAKNQAGOHNTDITLPSLRFHROWINGLGNPCSDTDLM 830
 766 YHSTGGINENARLFTVTP-EMFGAATGAPAHADALSILMLTRFADWVNALEKASSVLAA 824
 831 LRQOITLADRILASVMGLDISMVTQAMVSA-----GVNQLQCWODINTVLOWIDVA 880
 825 FEANSLTAEQLADAMNLDANILLQASIOAQNHQHLPPVTPENAFSCWTSINTILQWNVVA 884
 881 SALHTMSVIRTLNIRYVTAANKAESNLPSEDEWQTLAENWEAGLSTQOQAOTLADYTAE 940
 885 QQLAVAPQGVGALVGLDYIOSMKET---PTYAQWENAAAGVLTAGLNSQQANTLHAFUDE 940
 941 RLSSVLNCWFLANIQPEGVSLHSRDDLYSYFLIDNOVSSAIKTTRLAEAIAGIOLYINRA 1000
 941 SRSAAALSTYIRQAKAAALIKSRDDLYQYLLIDNOVSAALKTRIAEAIASIQLYYNRA 1000
 1001 LNRLEPNARADVSTRQFFTDW-TVNNRYSTWGGVSRVLYYPENYIDPTORIGOTRMMDEL 1059
 1001 LENVEENANSVISRQFPIDWDKYNKRYSTWAGVSQVLYYPENYIDPTMRIGOTKMDAL 1060
 1060 LENSQSLSRDTVEDAPKTYLTPETVADIKVVSAYHDNVNSLTGLTWFVQOTRENLUPE 1119
 1061 LQSVSQSLNADTVEDAFMSYLTSPFQVANLKVISAYHDNINNDQGLTYFTGLSETDAGE 1120
 1120 YYMNRNDISRMQAGELAANAWEKTKIDTAVNPKDAIRPVIPIERLHLIWEKEEVAK- 1178
 1121 YWKSVDHUSKENDGFAANANSEWHKIDCPINPKYSTIRPVIYKSRLLYLLWLEKEIKTQ 1180
 1179 --NGTD--PVETYDRFTLKLAFLRHGDSWSPWSDYITTOVEAVTDKPKDTERLALASG 1234
 1181 TGNKSGQYQETDVRVELKLAHRYDGTWNTPTITFDVNNKISL--KLEKNRAPGLYCAG 1238
 1235 FQGEDTLVYFKYTKGYSYSDGGSNKNVAGMTIYGDGSKFKMENTALSRYQLKNTFDII 1294
 1239 YQGEDTLVMPYNOQDTLDSY--KNASMOGLYIFADWASKDMTPQSNVYRD--NSYQOF 1294
 1295 HTQGNDLVRKASVYRPAQDFEVPASLNMGs--AIGDLSLTVMENGNIPOITSYSSDNLAI 1352
 1295 DTNN---VRRVNNRYAEDYEIPSSVSRKDYGWGDYILSMVYNGDIPINYKAAASDLKI 1351
 1353 -----TLHNAAFVRYDGSNGVIRNKQISAMKLTGVDGKSGQYGNAFIANTVRYGY 1405
 1352 YISPKLRIIHNG-----YEGQ---KRQCNLMNKG-----KLGDKEIV-----Y 1388
 1406 SDLG-GPITVYNTKQY-IASVQGHLMADYTRRLILTPVENNYVYARLFEF-PESPNTIL 1462
 1389 TSLGVNPNNSKNLMFYVYQYSGNTSGLNQGR--LLFHRDTPPYPSKVEAMIPGAKRSLT 1446
 1463 NTVPFTVGS-----NKTSDFKKCSYAVDGNNSQGFQIFSSYQSSGMLDIDTGINNTDIK 1515
 1447 NQNAIGDDYATDSLKNKPDLLKQYIFMTD---SKG---TATDVSGPVEINTAISPAKVQ 1499
 1516 ITVWAGSKTHFTASDHIASIPANSFDMPTTFKPLEIDASSLAFATNNIAPLDIVFETKA 1575
 1500 IIVRAGGKEQFTADKQVSIOFSPFDEMNQFNALEIDGSGLNFINNSASIDVTFTAPA 1559
 1576 KGRVILG---KIKOTLSKRVNYPEDILFLRETHSGAQYMOVGVRIRLNTLLASOLV 1631
 1560 EDGRKLGYESFSPVTLTKV-----STDNALTLHNENGAQIMQWOSYRIRLNTLTFARQLV 1614

Db 2373 NNLAFCAGTDTKTSIQASVSFADLKIREDPASLGKIRRIKOISVTLFPALLGPGYQDVQAI 2432
 QY 2457 LNYGSGVMPGCSAIALSHGVNDGQFMDLNDPNDKRYLPFFEGISVNDGSLTSPDPTAT- 2515
 Db 2433 LSYGDKAGLANCEALVSHGVNDGQFMDLNDKRYLPFFEGIAI-DQGTLTSPNASM 2491
 QY 2516 ---DRQKALLESLSIIHLIRYTR 2537
 Db 2492 PEKGQATWMLKTLNDIILHRYTIK 2516

RESULT 13

ID ADZ72157
 ADZ72157 standard; protein; 2516 AA.

XX AC ADZ72157;

XX DT 14-JUL-2005 (first entry)

XX DE P. luminescens toxin A protein SEQ ID NO:4.

XX KW transgenic plant; vector; pesticide resistance; insecticide resistance;
 herbicide resistance; insect resistance; toxin A.

XX OS Photorhabdus luminescens.

XX PN US2005102713-A1.

XX PD 12-MAY-2005.

XX PF 07-NOV-2003; 2003US-00703280.

XX PR 07-NOV-2003; 2003US-00703280.

XX PA (MERL/) MERLO D J.

XX PA (LIUD/) LIU D.

XX PA (BURT/) BURTON S L.

XX PA (GLAN/) GLANCY T P.

XX PI Merlo DJ, Liu D, Burton SL, Glancy TP;

XX DR WPI; 2005-384062/39.

XX CC New nucleic acid construct comprises a structural gene of interest
 PT functionally linked to untranslated regions (UTR), useful for enhancing
 PT expression of a recombinant nucleic acid sequence in transgenic plants
 PT and plant tissues.

XX PS Example; SEQ ID NO 4; 99pp; English.

XX CC The invention relates to a novel nucleic acid construct comprising at
 CC least one structural gene of interest functionally linked to one or more
 CC of an untranscribed region (UTR) selected from 3 fully defined 40-260
 CC base pair (bp) sequences (ADZ72154, ADZ72155 or ADZ72156), given in the
 CC specification or its derivatives or functional equivalents. Also claimed
 CC are a plant or plant cell transformed with the nucleic acid construct, a
 CC vector comprising the nucleic acid construct, a method for recombinantly
 CC producing a peptide or protein, a method of increasing expression of a
 CC gene in a plant cell, a recombinant DNA construct comprising a gene of
 CC interest and at least of a UTR selected from a 5' or a 3' UTR from an
 CC osmotin gene, a 5' UTR from an osmotin gene or its derivative
 CC functionally linked to a heterologous gene, a 3' UTR from an osmotin gene
 CC or its derivative functionally linked to a heterologous gene, and a
 CC recombinant nucleic acid construct comprising 5' and/or 3' UTR isolated
 CC or derived from an osmotin gene, the 5' and/or 3' UTR functionally linked
 CC to a structural gene of interest. In the nucleic acid construct, at least
 CC one structural gene of interest comprises a gene capable of conferring a
 CC non-native phenotype in a plant or an insecticide or herbicide resistance
 CC in a plant. At least one structural gene of interest also comprises an
 CC insect resistance gene isolated or derived from 5 fully defined 1750-7302
 CC bp sequences (ADZ72159-ADZ72163), given in the specification. At least
 CC one structural gene of interest further comprises a fully defined 7560 bp

CC sequence (ADZ72158), given in the specification or its derivative. The
 CC nucleic acid construct is useful for enhancing expression of a
 CC recombinant nucleic acid sequence in transgenic plants and plant tissues.
 CC The compositions and methods above are useful for modifying cells,
 CC tissues, or organisms using 5' or 3' UTR regions isolated or derived from
 CC a tobacco osmotin gene. The present sequence represents the P.
 CC luminescens toxin A protein, an insecticidal protein used in a vector of
 CC the invention.

XX SQ Sequence 2516 AA;

Query Match 38.7%; Score 5043; DB 9; Length 2516;

Best Local Similarity 42.6%; Pred. No. 0;

Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;

QY 22 LADQLYLSFSELRKIFDDQLSWGEGARHLYTHETIOKNNRLLEARIFTRANPQLSGARL 81
 Db 21 LTDISHSFNEFRQVSEHLSWSETHDYHDAQQAQKONRLYEARKRANPQLQNAVHL 80
 QY 82 GIERDSVSR-SYDEMFGARSSSVKPGSVAMFSPAGYLTETLYRAKDLHFSSAYHLDN 140
 Db 81 AILAPNAELIGVNNQFSGRASQYVAPGTVSSMFPAYLTETLYREARNLHSDSVYLT 140
 QY 141 RPDLDLTLQSNDTEISTLTLSNELLSHI-TRKTGSDSALMESISTYRQALDTPY 199
 Db 141 RRPDLKSMALSOQNMIDIELTSLUSNELLSIKTESKLENTKYVMELSTFRPGATPY 200
 QY 200 HOPVETIRQVIMTHDSTLSALSRNPEVVMQEGASLLAILANISPELYNLTETEEKNA 259
 Db 201 HDAYENVREVQLQDPGLEQLNASPAIAGLMHQASLLGINASISPELFNLTETEEGNA 260
 QY 260 DALPAQNFSENIPTENFASQSWIAKYGLESEVQKYLGMQLQN-GYSOSTSAIVYDNISGT 318
 Db 261 EELYKKNFG-NIEPASLAMPYELKRYNLSDEELSQFICKASNFGQOEYSN---NQLITP 316
 QY 319 LVVNNEKLEAYKITR-VKTDYDKNINYFDLMYEGNNOFFFRANFKVSRREGAT----- 372
 Db 317 VNSSDGTGVKYRITREYTTNAYQMDVELFP--FGGEN---YRLDYKFNFTNASTLSIK 371
 QY 373 -----LRKNAGPSGIVGSLGPLIANTFNPKSNLYNISDSEYKNGVKIYARYTTSPTS 425
 Db 372 LNDKRELVRTEGAQ-----VNIEYSANITLTAD--ISQPEIGLTRVLPSGS 418
 QY 426 ATNCGGIFTPESEYPLTIPALKKNAIRLCUTGSLSPNELQTVIRSDNAQGIINOSVLT 485
 Db 419 WA-YAAAKFTVEEYNQYSFLLKNAIRLSRATLSPTILEGIVRSVNLQDLNDIVLKG 477
 QY 486 VFYTLFYSHRYALSPDDAQLNGSVINQYADDSVSHENRLENTPLPKGKIFEADGNTVS 545
 Db 478 VFLTKYMQRYAIHAETALILCNAPISQSRSDNQPSPQDFRLFTPLNGQYFSTGDEEID 537
 QY 546 IDPDEEQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLTSVFVSISSLYRLTLAR 605
 Db 538 LN-SGSTGDWKTILKRAFNDIDVSLFRLLKITHDKDKGKKNLKNLSNLYIGKLLAD 596
 QY 606 VHQLTVNCLMYGLSPFNGKTTAS-LSSGBELPRVIMLYQVQMTLRAEITTEAILWLC 664
 Db 597 IHQTLIDELLL-IAGVEGKTNLSAISDKQLATLIRKLTNTITTSWLTKQKWSVOLFIMT 655
 QY 665 TPESGNTSPRISNLLNLRPSI-----SEDMAQSHNRLEQAEILAPPIAATLHASPDMAR 721
 Db 656 STSNKTLTPRIKLLDVTYHGLQGFDDKAD-----LLHVMABYIATLQLSSENVAH 709
 QY 722 YILLWTDNLRPGGLDIAGFWT-----LVLKEISLNANETIQ-----LVQFCHVMAQLSLS 770
 Db 710 SVLLWADKLQGD-----GAMTAERKFWDLNLTYPGSSSEAVETQEHIVQYCOALALENV 765
 QY 771 VQTLRLSBAELSVLVISGFVAVLGAKNPAGQHNIDTLFSLYRHFHWINGLGNPGSDTLD 830
 Db 766 YHSTGINENAFRLFVTKP-EMFGAATGAAPAHADALSILMLTRFADWVNALGEKASSVLAA 824
 QY 831 LRQQTILTRDLASVNGLDISWVTOAMVSA-----GVNQLQCQWDINTVLQWIDVA 880

Db 825 FEANSLTAEQLADAMNLDANLLQASIQAOHQHLPVPTPENAFSCWTSINTILQWVNA 884
Qy 881 SALHTMPSVIRTLVNIYVYTAALKAESNLPSWDEWOTLAENWEAGLSLTOQAQTADYAE 940
Db 885 QQLNVAPOGVALVGLDYIQSMKET-PTYAQWENAAAGVLTAGLNSQQANTLHAFLE 940
Qy 941 RLSSVLCNWFELANTIQEGVSLHSDRLYSYFLIDNOVSSAJTKTTLRAEAIAGIOLYNRA 1000
Db 941 SRSAALSTYIIRQVAKAAAKSRDLYLQYLLIDNOVSSAAIKTRIAEAIASIQYVNA 1000
Qy 1001 LNRPEPNARADYSTRQFTDW-TVNNRYSTWGGVSRVLYPENYIDPQRIGQTRMMDEL 1059
Db 1001 LENVEENANGVISHRQFFIDMDKYNKRYSTWAGVSQVLYPENYIDPTMRIGQTKMDAL 1060
Qy 1060 LENTISQSKSLRDTVEDAFKTYLTRPETVADILKVSAHYHDNVNSNGLTWFGQOTRENLPE 1119
Db 1061 LQSVSQSLNADTVEDAFMSYLTSEFQVANLKVTSAYHDNINNOGLTYFGLSETDAGE 1120
Qy 1120 YYWRNVDIRMQAGELAANAKWTKIDTAVNPYKDAIRPVI FRERLHLIWWKEEVAK- 1178
Db 1121 YYWRSVDHSHKFNDDGFAANANSEMHIKDCPINPKYSTIRPIYKSLRLLMLWLEQKEITYK 1180
Qy 1179 --NGTD--PVETDYRFTLKLAFRLHDSWSPMSYDITTOVEAVTDXKPDTERLALASG 1234
Db 1181 TGNKSGGYQTEDYRYELKLAHRYDGTWNTPIITFDVNKKISEL--KLEKNRAPGLYCAG 1238
Qy 1235 FQGEDTLVYVYTKSKSYDFGGSKNVAAGMTIYDGSFKOMENTALSRYSQLKNTFDII 1294
Db 1239 YQGEDTLVMPYNOQDTLDSY--KNASMOGLYIFADMAASKOMTPEQSNVYRD--NSYQOF 1294
Qy 1295 HTQGNDLVRKASYRPAQDFEYPASLNMGS--AIGDLSLTVMENGIPOITSKYSNDLAI 1352
Db 1295 DTNN---VRRVNNRYAEYIEIPSSVSRKDYGVGDYLLSWYNGDIPINTYKAASSDLKI 1351
Qy 1353 -----TLHNAFTVRYDGSNVIRNRKQISAMKLTGVGSKSOYGNAPFIANTVKHYGGY 1405
Db 1352 YISPKLRIIHNG-----YEGQ-----KRQCNLMNKG-----KLGDKEIV-----Y 1388
Qy 1406 SDLG-GPITVYNTKYN--IASVOGHLNADYTRRLILTPVNNYARLPEF-PPSPNTIL 1462
Db 1389 TSLGVNPNSSNKLMPYVYQYSGNTGLNQR--LLPHRDTPYPSKVEAWIPGAKRSLT 1446
Qy 1463 NTVFTVGS-----NKTSDFKCSYADVGNNSQGFQIFSSVQSGSLDIDTGNNTDIK 1515
Db 1447 NONAAGDDYATDLSKEDDLKQYIFMTD-----SKG-----TADVSGPVEINTALSPAKV 1499
Qy 1516 ITVMAGSKTHRTASDHIALSPANSFDAMPYTFKPLEIDASSLAFTNNIAPLDIVPETKA 1575
Db 1500 IIVKAGGKEQFTADKDVSIQPSRPSFDEMNYQFNALEIDGSGLNFINNSASIDVTFTAF 1559
Qy 1576 KGRVIG-----KIKOTLSVKRVNVPEDILFRETSHGAQYMQVLGVYRIRLNTLASOLV 1631
Db 1560 EDGERKLGYESFSIPVTLKV-----STDNALTLHHNENGAQYMQVSYRTRLTLPAROLV 1614
Qy 1632 SRANTGIDTILTMETQRLPEPLGGRFANFVLPKYDPAEHGDSRWFKIHGNVCGNTR 1691
Db 1615 ARATGIDTILISMETQNLQEPOLGKGFYATVPYPLNSTHGDERNFKLYIKHVVDNNS- 1673
Qy 1692 QPYTSGMLSDTSETSMTLF-----VPIAEGYMHGVRGLGVYQKIITYDNTWESAFFYD 1746
Db 1674 HIIYSGQLTDIN-INITLFIPLDDVPLNQDYH---AKVYMTFKKSPSDGTWGWGPHFVRD 1728
Qy 1747 ETQOQFVLINADHDSGMTQOQIVKNYIKKYKGLFNLVSIATGYSA PMDFNSASALYYWELF 1806
Db 1729 D--KGIVTIN-----PKSILTHFESVNVLNIS-----SEPMDFSGANSLYFWELF 1772
Qy 1807 YYPWMCQFRLLOEKQFDEATQWVNVNYPAGYVINGBIAPIWNCRPLEETTSNANPL 1866
Db 1773 YITPMLVAQRLHEQNFDEANRWLYKTVWSPSGYIVHGQIQNYQMNVRPLEEDTSWNSDPL 1832
Qy 1867 DAIDPDVAQNDDPMHYKIATFMRLLDQILRGDMAYRELTRDALNEAKOWYVTRLELGD 1926
Db 1833 DSVDPDAVAQDPMHYKYSTFWRTLDDLLIARGDHAYRQLERDTLNEASWYMQALHLLGD 1892

Qy 1927 EPEDYGSQWAAAPSISLGAASQTVQAAQODLTMLGRG--GVSKNLRTRANSVLGLFLPEY 1983
Db 1993 KPYPPLSTWSDPRDRAADITTONAHDSAIVALRQNIPTPAPLSLSANVTDLFLQOI 1952
Qy 1984 NPALTDYQWTLRLRLFNLRHNLSDIGQPLSLAIYAEPDTPKALLTSMVQASQGSASVLP 2043
Db 1953 NEVMNMYWQTLAQRVYNLRHNLSDIGQPLYPYATPADPKALLSAAVATSGGKLPES 2012
Qy 2044 TLSLYRFPVMLERTNLVAQLTQGTSLLSMAEHDDADELTLLLOQGMELATOSIRQQ 2103
Db 2013 FMSLWRFPMLLENARGMVSQLTQFGSTLQNIERQDAEALNALLQNAEILNLISQD 2072
Qy 2104 RTVDEVDADIATLAESRSRAQRNLEKYQOLYDEINHEGEORAMSLDAAAGOSLAGVLS 2163
Db 2073 KTIIBELDAEKTVLEKSKAGASRFDYSKGLYDENINAGENQAMTIRASAGLTTAVQASR 2132
Qy 2164 IAEGVADILVNPVFLGACGSRGWAALRASASVMSLSATASQYSADKISRSBAYRRRQEW 2223
Db 2133 LAGAADILVNPVFGAGGSRWGAIAEATGVMEFSANVMNTEADKISQSETYRRRRQEW 2192
Qy 2224 BIQRNADGEVKQMDAQLESKIRREAAQOVYQETOQAHTQAOLELLQRKFTNKALYS 2283
Db 2193 EIQRNNAEELKQDAQLKSLAVREAAVLQKTSKTQEQTSQSLAFLOKRFKFSNOALYN 2252
Qy 2284 WMRGKLSAIYYQFDLTQSFCLMAQEALRRELTONGVTFIRGGAWNGTTAGLMAGETLLL 2343
Db 2253 WLRGLAAIYQFQYDLAVARCLMAEQAYRWELNDDSRFIPGAWQGTYAGLLAGETLML 2312
Qy 2344 NLAEMEKVLERDERALEVTRTVSLAQFYQALSSDN--FNLTEKLTQFLREGKGNVGSAG 2401
Db 2313 SLAQMEDAHLKRDKRALEVERTVSLAEVYAGLPKDNQGPFLSAQEBIDKLVSQSGSAGSGN 2372
Qy 2402 NELKL-----SNRQIBASVRLSDLKIFSDYPESLGNTRQLKOVSVTLPALVGPYEDIPAV 2456
Db 2373 NNLAFGAGTDTKTSLOASVSPADLKIREDPASLSGKIRRIKQISVTLPALGYPQDVQAI 2432
Qy 2457 LNYGSGIVMPRGCSAIALSHGVNDSGOFMLDFNDSRYLPPFEGISVNDSGSLTLGFPDAT- 2515
Db 2433 LSYGKGAGLNGCBALAVSHGVNDSGQFQDFNDGKFLPFEGIAI-DQGTTLTLSPFNASM 2491
Qy 2516 ---DRQKALLESLSDIILHRYTIR 2537
Db 2492 PEKKGQATMLKTLNDIILHRYTIK 2516

RESULT 14
ID AAB72611 standard; protein; 2517 AA.
XX AAB72611;
AC AAB72611;
XX 04-MAY-2001 (first entry)
XX Modified Photorhabdus tcdA toxin.
XX TcdA; TcdA; insect toxin; plant; insect resistance.
XX Photorhabdus sp.
OS Synthetic.
XX WO200111029-A1.
XX 15-FEB-2001.
PD 11-AUG-2000; 2000WO-US022237.
PF 11-AUG-1999; 99US-0148356P.
PR (DOWC) DOW AGROSCIENCES LLC.
PA Petell JK, Merlo DJ, Herman RA, Roberts JL, Guo L, Schafer BW,
PI Sukhapinda K, Merlo AO;

[illegible]

Db 559 LN-SGSTGDRKTIILKRAFNIDDDVSLFRLLKTTIDHNDKDKGKIKNNLKNLSNLIYIGKLLAD 617
Qy 606 VHLQTVNELCMLYGLSPFNKKTAS--LSSGELPRVIMLYQVOTOMLTAEBEITTEAIWLIC 664
Db 618 IHLQIIDEIDLL--IAGBGKTNLSAISDKQLATLIRKLTITSMHTQKWSVFLPMT 676
Qy 665 TPEFGSNTSPEISNLLNLRPSI-----SEMAOSHRELOAEILAPFIANTLHLPDMAR 721
Db 677 STSYNKLTPETIKNLLDTPYHGLQDFDKAD-----LLHVMAPYIAATLQSSENVAH 730
Qy 722 YLLMTDNLRLPGGLDIAGPMT-----LVAKESLNANETQ-----LVQFCHVMAQLSLS 770
Db 731 SVLLWADKLQPD-----GAWTAEKFWDLNLTYPGSSSEAVTQEHVOYICQALQLEWV 786
Qy 771 VOTLRSEAEBSLVISGPAVLGAKNQAGQHNIDTFLSLYFHOWINGLNGPFGSDTIDM 830
Db 787 YHSTGINENAPRLFVTKP--EMFGAATGAAPAHADALSILMTFADWVNALGKASSVLAA 845
Qy 831 LRQQTITADRSLASVGLDISMTQAMVSA-----GVNLOCWQDQINTVLOWIDVA 880
Db 846 FEANSILTAEQADAMNLDANLLLOASIOAQNHOHLPPVTPPENAFSCWTSINTILOQWVA 905
Qy 881 SALHTMPSVIRTLNIRVYVYALNKAESNLPSWDEWOTLAENMEAGLSTCOAOOTLADYTA 940
Db 906 QQLNVAPOGVSALVDLYIQSMKET-----PTTAQWENAGVLTAGLNSQOANTLHAFIDE 961
Qy 941 RLSSVLCNWFANIQPEGVLSHSDLYSYFLIDNQVSSAIKTRRLABAIAGIQLYINRA 1000
Db 962 SESAALSTYIYIQAQAAKAAIKSRDLYQLLIDNQVSSAIKTRIAEALIASIQLYVNA 1021
Qy 1001 LNRIENARAADVSTQPFDM--TVNRYSTWGVSRVLYPPENYIDPTORIGQTRMDEL 1059
Db 1022 LENVENANSGVISRFQFIDMDKYNKYSTWAGVSLVYPPENYIDPTWIRIGQTRMDAL 1081
Qy 1060 LENISQSLRDTVDPAKTYLTRFETVADLVKVSAYHDNVNSNGLTFWVGQTRLENPE 1119
Db 1082 LQSVSQSLNADTVEDAFMSYLTSEQVANLKVISAYHDNINNDQGLTYFGLSETDAGE 1141
Qy 1120 YVRNVDIRMQAGELANAKWETKIDTAVNPKYDAIRPVIFRERLHLIWEKEEVAK- 1178
Db 1142 YVRSVDHSKFNDFKFAANANSEWHKIDCPINPKYSTIRPVYKSRVLYLMLBQKEITKQ 1201
Qy 1179 --NGTD--PVETDYRFTKLAPLRHDGWSAPWSYDITTVQEAUVTDKPDTERLALAA 1234
Db 1202 TGNKSGQYQETDYRYELKLAHRYDGTWNTPTFDVNNKISEL--KLEKNRAGPLYCAG 1259
Qy 1235 FQGEDTLVYVYKTKGYSDFGSKNKNVAGMTIYDGSFKKMENTALSRYSLKNTFDII 1294
Db 1260 YQGEDTLVYVYNOQDITLDSY--KNASMOGLYIFADMASKONTPEQSNVYRD--NSYQOF 1315
Qy 1295 HTQGNDLVRKASYRPAQPEVPASLNMGS--AIGDSDLTVMENGNIPQITTSYSSDNLAI 1352
Db 1316 DTNN--VRRVNNRYAEDEYIEPSSVSRKDYGMGYLLSMVYNGDIPITNYKAASSDLKI 1372
Qy 1353 -----TLHNAFTVRYDGSNGVTRNKQISAMKLTGVDGKSGYQGNAPFIANTVRYGGY 1405
Db 1373 YISPKLRIRHNG-----YEGQ----KRNQCNLMNKG-----KLGDKEIV-----Y 1409
Qy 1406 SDLG--GPITVYKTKNY--IASVQGHLMADYTRRLILTPVENNYARLPEF--PSPNTIL 1462
Db 1410 TSLGVNNNSNKLMPYVYQYSGNTSGLNQGR--LLFHRDTPYPSKVEAMIPGAKRSLT 1467
Qy 1463 NTFTVGS-----NKTSDFKCSYAVDGNNSQGFQIFSSYQSGWLDIDTGINNTDIK 1515
Db 1468 NQRAAIGDDYATDLSLNKPDLLKOYLFMTD--SKG-----TATDVSGPVEINTAISPQVQ 1520
Qy 1516 ITWAGSKTHTTASDHIAISPANSFDMPTTFKPLEIDASSLAFTNNTIAPLDIYFETKA 1575
Db 1521 IIVKAGGKEQTTADKQVSIQSPSPFDEMNYQFNALEIDGSGLNFINNASSIDVFTTAPA 1580
Qy 1576 KQGRVLG-----KIKOTLSVKRVNNYPEDILFLRETHSGAQYMQLVGYRIRLNTLASQIV 1631
Db 1581 EDGRKLGYESFIPVTLKY-----STDNALTLHNNENGAQYMQWOSYRTRLNTLFAQLV 1635

Qy 1632 SRANTGIDITILMETQRLPEPPLGEGFPANFVLPKYDPAEHGDERWFKIHIGNVGNTGR 1691
Db 1636 ARATGIDITILSMETQIQEPOLGKGFATFVIPPYNLSTHGERWFKLYIKHVYDNNNS- 1694
Qy 1692 QPYYSGLMSDPTSEVSMTLF-----VPYAEGYMHGEGVRLGVGYQKITYDNTWESAPFYD 1746
Db 1695 HIIYSQJTDN-INITILFILDVPLNODVH-----AKYMTFFKSPSDGTWNGPHFVRD 1749
Qy 1747 ETQOQVILINDADHSDGTMTOQGIKNIKKYKGLNVSITATGYSAFMDFNASALYWFEL 1806
Db 1750 D--KGIIVIN-----PKSILTFESVNVNINIS-----SEPMDFSGANSLYFWEL 1793
Qy 1807 YITPMFCQRLLOEKQFDEATOWINYYNYPAGYINVEIAPWNCRLPEETTSNNANPL 1866
Db 1794 YITPMVLAQRLHLEQNFDEANRWLYWVSPSGYIYHGOITQYQMVNRPLLEDTSWNSDL 1853
Qy 1867 DAIDPDVAQNDPMHYKIATFMRLLDOLILRGDMAYRELTRDALNEAKWYVYRTLELGD 1926
Db 1854 DSVDPDAVAQHDPMHYKYVSTFMRLLDOLILRGDMAYRELTRDALNEAKWYVYRTLELGD 1913
Qy 1927 EPEDYGSQQAAPSLSGAASQTVQAAQODLTMLGRG---GVSKNLTANSVLGLFLPEY 1983
Db 1914 KPYLPSTWSDPRLDRAADITTQNAHDSAIVALRQNIPTPAPLSRSANTITDLFLPOI 1973
Qy 1984 NPALTDYQTLRLRLFNLRHNLSDIGQPLSLAIYAEPTDPKALLTSMVOASQGSVALPG 2043
Db 1974 NEVMVNYQTLAQRYVNLRHNLSDIGQPLYLPIYATPADPKALLSAAVATSQGGKLPES 2033
Qy 2044 TLSLYRFPVLMERTNLVAQLTOGTSLLSMAEHDADDELITLLLOQGMELATQISIRQO 2103
Db 2034 FMSLWRFPHMLNARGMVSQLTOFGSTLQNTIERODAEALNALLQNAEGLITLNLSD 2093
Qy 2104 RTVEVDADIIVLAESRRSAQNRLEKYQOLYDEIDNHGEQRAMSLDAAAGOSLAGQVLS 2163
Db 2094 KTIEELDAEKTVEKSKAGASRPDSYKLYDENINAGENQAMTLASAAGLITAVQASR 2153
Qy 2164 IAGVADIVPNVFLGACGSGRWGAALRASAVMSLSATASOVSADKISRSEAYRRRRQEW 2223
Db 2154 LAGAAADLVPIIFGAGGSRWGAATAEATGYVMEFSANVMNTADKISRSEAYRRRRQEW 2213
Qy 2224 EIORNDAGEVKQMDAOLESIKIRREAAQMVVEYOETQOAHQAOLELLORFKFTNKALYS 2283
Db 2214 EIORNNAEELQIDAQKSLAVRREAAVLQKTSUKTOOQTSQQLAFQKFSNOALYN 2273
Qy 2284 WMRGKLSAIYYQFDPDLTOSFCLMAQEARRELTONGVTFIRGAWNGTTAGLMAGETLIL 2343
Db 2274 WIRGELAIIYQFYDLAVARCLMAEQAYRWELNDDSAFIFPGAWQGYIAGLLAGETLML 2333
Qy 2344 NLAEMEKVLMRDERALEVTRTVSLAQFYQALSSDN--FNLTEKLTQFLREKGNVAGSG 2401
Db 2334 SLAQMEDAHLKRDKRALEVERTVSLAEVYAGLPKONGPFSLAQEDKLVSOQSGSAGSN 2393
Qy 2402 NELKL-----SNRQIEASVRLSDKI FSDYPESLGNTRQLKOVSVTLPALVGPYSDIRAV 2456
Db 2394 NNLAFAGACTDTKTSLQASVSFADLKIRIDEPASLGKIRIKOISVTLFALLGYPQDOQAI 2453
Qy 2457 LNYGSSIVMPRGCSAIALSHGVNDSSQGMFLDNFSRYPFEGISVNDSSGSLTSPFDAT- 2515
Db 2454 LSYGDKAGLANGCEALAVSHGWNDSGQOLFQDNQKFLPFEGIAI--DQGTLTLSFPNASM 2512
Qy 2516 ---DRQKALLESIDIIILHRYTIR 2537
Db 2513 PEKGQAATMLATLNDIILHRYTIR 2537

Search completed: February 16, 2006, 21:31:20
Job time : 263.656 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:31:49 ; Search time 56.6609 Seconds
(without alignments)
4309.818 Million cell updates/sec

Title: US-10-754-115-34

Perfect score: 13043

Sequence: 1 MYSTAVLLNKISPTRDQTM.....KALLESLSDIILHRYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_80.*

1: PIR1.*

2: PIR2.*

3: PIR3.*

4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	317	2.4	833	2 AB0448	probable insectici
2	254.5	2.0	4688	2 F92885	hypothetical prote
3	240.5	1.8	2893	2 A64556	toxin-like outer m
4	235	1.8	2178	2 S55805	alpha-toxin - Cios
5	229.5	1.8	2364	2 I40884	cytotoxin L - Cios
6	229.5	1.8	2383	2 D64962	probable membrane
7	222	1.7	2710	2 A37052	toxin A - Clostrid
8	222	1.7	5291	2 F90696	hypothetical prote
9	221.5	1.7	2902	2 C71953	toxin-like outer m
10	220	1.7	5005	2 F62884	hypothetical prote
11	220	1.7	5188	2 B85547	probable RTX famil
12	217	1.7	1830	2 E92909	conserved hypothet
13	217	1.7	2628	2 S59413	probable membrane
14	216	1.7	2529	2 B64635	toxin-like outer m
15	216	1.7	4385	2 T29042	hypothetical prote
16	211	1.6	2819	2 A30551	conserved hypothet
17	211	1.6	6713	2 B89921	hypothetical prote
18	205	1.6	1856	2 C95008	immunoglobulin A1
19	203.5	1.6	2491	2 A57036	talin - elime mold
20	202.5	1.6	3194	2 D71917	toxin-like outer m
21	201.5	1.5	2367	2 S70172	toxin B - Clostrid
22	201	1.5	1819	2 D97033	uncharacterized pr
23	200.5	1.5	1795	2 F97713	190K antigen precu
24	200	1.5	1628	2 E90538	hypothetical prote
25	199	1.5	2660	2 E85822	probable invasin Z
26	196	1.5	4199	2 S76412	hypothetical prote
27	195	1.5	3705	2 A00123	probable autotrans
28	194	1.5	1962	2 A32634	lactocepin (SC 3.4
29	194	1.5	3013	2 AB0480	probable invasin Y

30 193.5 1.5 2052 2 C97038 phage-related prot
31 192.5 1.5 1965 2 S75200 fat protein - Syne
32 192.5 1.5 3216 2 C90538 hypothetical prote
33 192.5 1.5 3283 2 AC1018 large repetitive p
34 192 1.5 2167 2 AF1489 cell wall-associat
35 192 1.5 2302 2 T14328 protein-tyrosine-p
36 191.5 1.5 2399 2 H71879 toxin-like outer m
37 190.5 1.5 1808 2 AB1847 serine/threonine k
38 190.5 1.5 4725 1 A44357 dynein heavy chain
39 189 1.4 1612 2 AB1347 probable peptidogl
40 188.5 1.4 1939 2 D97316 probable S-layer p
41 188.5 1.4 2059 2 D82671 surface protein xP
42 188.5 1.4 2475 2 T00047 gellan lyase (EC 4
43 188 1.4 1561 2 S61314 IGA-specific metal
44 188 1.4 3351 2 T13812 lipophorin - fruit
45 188 1.4 4936 2 AH2515 hypothetical prote

ALIGNMENTS

RESULT 1

AB0448

Probable insecticidal toxin YPO3681 [imported] - Yersinia pestis (strain CO92)

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004

C:Accession: AB0448

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

ll, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,

Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AB0448

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-833 <R>

A:Cross-references: UNIPROT:Q8ZAV3; UNIPARC:UPI000000CDA68; GB:AL590842; PIDN:CAC93150.1;

C:Genetics:

A:Gene: YPO3681

Query Match 2.4%; Score 317; DB 2; Length 833;

Best Local Similarity 18.8%; Pred. No. 4.9e-09;

Matches 179; Conservative 151; Mismatches 314; Indels 310; Gaps 31;

Qy 49 LYHETIEQKNRLLLEARIIFTRANPOLSGAURLGIE-----RDSVRSYDEMFGARS 100
Db 19 IHQAKAEERHQAQVRATQLRNDPVLGINKLAFSVAPKILQPEARLDSLAEGIPERA 78
Qy 101 SSFVKPGSVASMFSPAGYLTLYREAKDLHFSSAYHLDNRRPDLADLTLSQSNMDTIS 160
Db 79 NEYADPASIQSLSPGRLCYLXYHAKELHEDGNKHLIDKRRPDLQELVLSNNMQEVS 138
Qy 161 TLTSNLLLEHITRTKGGSDALMESLS--TYRQADTPYHQPYETIRQVIMTHDSTLS 218
Db 139 SL-----BILLNVL-----QTNAPLAKLAKDTEAHANDVSFTLPY-----DDNLT 178
Qy 219 ALSRNPVWGQAGASLLAILANISPELYNLTITEIKNADALFAON---FSEMITPEN 275
Db 179 VIN-----AILED-----KAISURETAALLAENNDPWANPITP-- 211
Qy 276 PASQSWIAKYGLBELSEVQKYLGMQLONGYSDSTSAYVDNIISTGLVNNESKLEAYKITRV 335
Db 212 -----ALVQEQGLNPASYA-----LIDIKSP----- 234
Qy 336 KTDYDKNNINFDLMYEGNNOFFIRANPKVSREFGATURKNAGPSGIVGSLSGPLIANTN 395
Db 235 --DD----- 236
Qy 396 FKSNNLSNISDSEYKNGVKIYAYRYTSTSATNOGGGIFTFESYPLTIFALKLNKAIRLC 455
Db 237 -----NSAKRLA 243

Db 2302 TFFTEIINDNNKTFPFNKSDSIQDKFTVLNNQVGVGNIIEIQDRDNNHLSAKIRFEL 2361
Qy 1060 --LENISQSLSDTVEPAFTYLRFPETADLVKVS---AYHDNVNSNTGLTWFGQT 1113
Db 2362 NDLNV---LSND--EQATTYNNQTTSA--KVITDQOKYLEATFSLVLM-----2407
Qy 1114 RENLPEYWRVNDISRMQAGLAANAKWTKIDTAVNPYKDAIRPVIFRE-----RLHL 1168
Db 2408 -----KDTIINKIEFNTKPNASKNIGINDT--NVIYDATNLINNDLKITGLHT 2456
Qy 1169 IWVEKEEVAKGTD-----PVETVDRFTLKLAPL-----RHGS--WSAPWYD--ITTOVE 1216
Db 2457 L---KEPEANNKTNISVSELDTNHHSKLNLYFAKFDSDGQSVLTNPAAADKIVTNNN 2513
Qy 1217 AVTDKPPTEBLAALASG---FOGEDTLVVFVYTKGYSYDFGG---SNKNV-----1262
Db 2514 --NNKKELTFTNLNLTNSRQYIFKG---LYVNSNGTNDENNKEKNSVDYKIVK 2567
Qy 1263 -AGMTIYDGSF-----KMENTALSRYSQLKNTFDIHH-----TQGNDLVR 1303
Db 2568 PTITTIQKNGNWTPOPNAQKFKFNINSDENVDFSTDLDAITTFSDQHDTKTITTKLKQ 2627
Qy 1304 KASTRPAQDFEVPASLNGSAIGDSDSLTWENG-----NIPQI---TSKYSSDWLA 1351
Db 2628 KDNQWYIEDTISDLAYNDTYKLDISITKPMNAPANLKIOTENKEQISFTTQSGPTQLVS 2687
Qy 1352 ILLHNAFTVRYDGSNNVIRNKOISAMKLTGVDGKSQYGNAPFIANTVKHYGGYSDL---1408
Db 2688 ITSDHLMNDQFNAS-----NQQTITAKVGVN--DLYNNRKI--KLVEYRSNNQKSLV 2738
Qy 1409 -GGPITVYNTKNTYASVQGHLMNADYTRRLILTPVENNYVARLFEPPFSPNTILNTVFT 1467
Db 2739 ESNELTQKQTOYIFLPISVANRQYSFKEIKITSNNPFEL-----NNSTNVTSFN 2793
Qy 1468 VGSNKTDFKCSYAVDGNQSGRQIFSSYOSSGWLDIDTGINNVDIKITWAGSKHTFP 1527
Db 2794 VNASKT-----QIVIDNTN-----LATNITSTAITTYKLAKSKOHVF 2830
Qy 1528 TASHIASLPANSPDAMPYTFKPLEIDASSLAFTNNIAPLD-----IVFET---KAK 1576
Db 2831 QVGSIIITLKSND-----ERQESYTKTITSISDGSSEATVFNTOQLKEE 2879
Qy 1577 DGRVLGKI---KQTLVKRVNYNPEDILFLRETHSGAQYMQL-----1615
Db 2880 ANYKLIKGVFKPFLATYNINDANNVIF--EDNNSNVNFKTLIVDHKVTNVSSNDSTINT 2938
Qy 1616 -----GVYRIRLNTLASQLVSRANTGIDTILTMETQRLPBPPLGEGFPANFVLP 1665
Db 2939 TTQTVNIDIDGIQRTWINKKI--QIVYTSNDGEEILSDQKTL-----LWAN---2982
Qy 1666 KYDPAEHGDERWFKHIGNVGNTGRQPYYSGLSDTSETSMTLFVFPYABGYMHGVR- 1724
Db 2983 -----NHYSFELNKHNRKYLUKEVRIINDNNKTSII-----FHLKNGIAD 3024
Qy 1725 --LGVGYQKITVDNTWE---SAFFYFDETKQFVLINDADHDSGMTOQGIKVNKKYKGF 1779
Db 3025 WPIVNTQSISSIIIEPIAKKNLQSTQIRFIL--NDPDNVLNNEEBAI-----3073
Qy 1780 LNVSIATGYSAPMDFNSASALYWEFLFYTPMPCFQRLLOKQKQDEATQWNYVNPAGY 1839
Db 3074 --INVGENLSVKAKVKVIGSKQYLE-----TTFNNLALNQ--DTIINSITFINPKPIKA 3122
Qy 1840 IVNGEIAPIWNCRLPBEETSWANPL-----DAIDP-----1871
Db 3123 ATNIGID-----NSNKIYNNNTN--NINPLKIDNNFSVIGPVQSDTKRITLSEYNNKTNVKUN 3177
Qy 1872 -----DAVAQNDDPMHYKIATFMRLLDQILIRGDMAYRELTRDALNEAKMYVRLLELLG 1925
Db 3178 LELKANDNILQN---LYFIATFKQNNETVAKSIVSYSSIKKEGKQ-----IIEI--3225
Qy 1926 DEPEDYGSQWAAAPSLSGAASQTVQAAAY-----QODLTWLRGGGVSKNLRNTANSILVGLFLP 1981

Db 3226 -----RIPNLKSNRLYHFDNLYYLLNQDDKOLIDNNKLNIRSNVNEI-----3268
Qy 1982 EYNPALTDY-----WOTLRLEFNLRHNLNSIDQOPLSLAIYABPTDPKALLTSMVQA--2033
Db 3269 EVKFGNTTFKKNVEMQ-----INSNSVNFKLQVB--SDDEDILDNNVMADV 3313
Qy 2034 -----SQGSASVLPGLTSLYR-----FPVMLETRNLVAQLTQ 2066
Db 3314 IFSSKSNQNDTKFVNNVNLKKEGNIWFIEGIVNNLKPETEYELKSIVLTKPLNANSNLT 3373
Qy 2067 FGTSLLSMAEHD-----DADELTTLLQQQEMELATQSIRIQ-----QRTVDEV 2109
Db 3374 INPTNISFKYKPGNGYIINIESNOTLTNSQNVNTIDGIRSEWPNKEAKLVISNTNDEV 3433
Qy 2110 DADIATLAESRRSAQNRL-----KYQOQLYD-----2135
Db 3434 EAKVNL-----BENKLVYVFNFNKLNKYNRKYTKSIKWPFGQNWIDFPNKNQNLQNSFI 3486
Qy 2136 --EDINHEGQRAMSLDDAAAGQ--SL--AGQVLSIAEGVADLVPNVFLGACGGS 2183
Db 3487 IKGKTPISVNNKLANEISDRVINNLNVAKIQFSLNDPEDVLSNNE--IAEIT-----3534
Qy 2184 RWGAALRASASV-----MSLSATASOVSADK--ISRSEAYRRRRQSEW---IORDNAD 2231
Db 3535 -YDNTKKTNAKVITVNNKYLEVLSNLEFNNHDTIINKIEFITRPNKAVDSVGINNSKI 3593
Qy 2232 GEVKOMDAQLESKIR-----REAAQWQVEYQFTQQAHTQAO 2268
Db 3594 YELTNLIBETRKAKINNDFRISGPLSSNNLHLEIKYNDKESINSSIEF--TINKHIAKN 3651
Qy 2269 LELLQRFPTN-----KALYSWWRGKLSAIYY-----O 2295
Db 3652 LK-FKLFRFBFGHEVYSSELIDGNINNKTKQEVLFVSVKGLSKNQLYKLVVYYLDNTHQ 3710
Qy 2296 PFDLTQSFCLMAQALARELTDN--GVTFI--RGGAWNGTTAGLMAGETLLLLAEME--KVM 2352
Db 3711 NIDTRK--IPKHNVSKHEIBINFGVTWISKHGNKSPD-----TTANPEFKIE 3758
Qy 2353 LERDERAL-EVYRTVTSLAQFYQALSSDNFNUTEKLTQFLREG-----KGNV GASNEUKL 2406
Db 3759 TQDDNDVLANNIDAIVK-----FKDEHNNIKQIVRIKENNDWLKIQDNLNLPETKY 3811
Qy 2407 SNROIE-----ASVRLSD-----LKIFSDYPESLCNRQLKQSVT 2442
Db 3812 KLENIELSKPLKTHLSVINDKENISLITETGNPVLKVIQTQNDTINDTQO--TINVT 3869
Qy 2443 L.PALVGVP--EDIRAVLNYGSGIYVMPRGCSAIALSHGVNDSGQFMDPNDKRYLPFEGIS 2500
Db 3870 LSGVNSKINGRQIKVVYKDNNNVIYE--SSLITLQKKNQYQLLLSNLSNREYRFEKIE 3927
Qy 2501 VNDSGSLTSPDPDATDRQKALLLESLSDI 2528
Db 3928 INHI-SNTNNPFD-----LEKLVG 3946

RESULT 3

A64556

toxin-like outer membrane protein Hp0289 - Helicobacter pylori (strain 26695)

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 09-Jul-2004

C:Accession: A64556

R:Tomb, J.F.; White, O.; Kervavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.

Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenne

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.

Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467; PMID:9252185

A:Accession: A64556

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-2893 <TOM>

A:Cross-references: UNIPROT:Q25063; UNIPARC:UPI000000D3097; GB:AE000547; GB:AE000511; NID

Query Match		1.8%; Score 240.5; DB 2; Length 2893;
Best Local Similarity		18.6%; Pred. No. 0.0067;
Matches 361; Conservative 275; Mismatches 745; Indels 563; Gaps 79;		
Qy	56	QKQNRRLARIFTRANPQIS-GAIRGIERDSVRSYDEMFGARSSFFVKPGSVASMF5 114
Db	1234	QAQNSYFIDDTAQQGNGNFESFALNL-...DFSNSFRGVGTQSVFKFNAKNAISFT 1289
Qy	115	PAGVITE-LYR-EAKDLHFSSAYHLNRRDLADLTLSQSNMDTEISTLTSLNELL-- 170
Db	1290	NSTNLSSGLYQMAQSFLFDS-...NLSVSVGTSSIKANAINLSQASINA 1337
Qy	171	-EHTTRKTGSDAL-MESISTYROAIDTPYHQPVTIROVIMTHDSTLS-...ALSNP 225
Db	1338	SNHSTLEQLDNLVNDTSSLMNOSTIVNSNATINDYASLIANGSHLNGAVNFNS- 1396
Qy	226	VMGOAGASLAILANISPELYNLTBEITEKNADALFAQ-NFSEN-...ITP- 273
Db	1397	-----ANITTSLN-...SSIIVFGAVSLGGQFNLSNNSLDFQGSALTIN 1439
Qy	274	-----ENFASQWIAKYGLESEVOKYLGMLQNGSDSTSATVNMISTGLVNNESKL 327
Db	1440	TAFFYDNFASQSPITTHQALDIKAPLSLGGNLIN-...PNNSSVLDLKNQSLVFGDQSL 1496
Qy	328	EAYKITRVKTDYDKNINYPDLMEGNNQPPIRANEKVSREFGATLR-...KNAGPSG 381
Db	1497	NIANIDLSDLNKNRNVYNIQADMNSWYTERISF-...FGHIINDGIYDAKNQYIS- 1550
Qy	382	IVGSLSGPLANTNFKSNYLS-...NISDSFYKNGKVIYAYRTSSTSATNQGGIF 434
Db	1551	FTNPLNALKITESFKDNQSLSVLSQIPGINKTLYNIGSEIFNYQ-...KVYNNANGVY 1605
Qy	435	TFESYPIITIPALKUNKAIRCLTSLSPNELQTVRSNDAQIINDSVILKVFYTLFYSH 494
Db	1606	SY-...SDDAQGVFLTSNKGYY-... 1625
Qy	495	RYALSFDQAQVLINGSVINOVADDSVSHENRLENTPLPKGIPEADGNTVSIIDPDEQST 554
Db	1626	-----NPNQSVQSGSNNTTKNNLTSESSIIISQTYNAQNPIS-... 1664
Qy	555	PARSALMRGLGVNGELYQGLAGVLDQAQ-NTITLSVFVISSLYRLTLARHOLT-... 610
Db	1665	---ALHIYNGYNFNKIALQMAKALKYPEIKKVLGNDFFSSLNALSNA-LNQLKLI 1720
Qy	611	-----VNELCMYLGLSPFNKKTASLSGSELPRVLWLVQVQWLTAEAITTEALWLL 663
Db	1721	TPNDWKINEL-...IDNANNVYQNFNG-...TLIVGATQIGQDTNSAVVFGGLGYQ 1772
Qy	664	CTPEFSGNISPEI-SNLLNLRPSISEDMAQSHNRELAQETLAPFAATLHLASP-... 717
Db	1773	TPCDYDIDVQKFRGTGLQGLLESSADLGYDITTFNAKEI-...YLTGTLGSGNAGTGG 1829
Qy	718	-----DMARYILLWTNLRPGGLDIAGFMFLVLKESLN-...ANETQL-VQFCH 762
Db	1830	SASVTPNSQTSLLIINQANIVSSQTD--GIFSMLOQEGINKVFNQAGLANILGEVAVQSI 1887
Qy	763	VMAQL-SLSVOTLRLSBAELSVLVISGPAVLGAKNP-...AGQHNIDTFLSLYRFHOWI 817
Db	1888	KAGGLGNLIVNTLG-...SNSVIGVYLTPEQKNQTLSQLGQNNFDNL-... 1931
Qy	818	NGLNPGSDTL--DMLRQ-...TLTADRLASVGLDLSMVTQAMVSAGVNOQCWODI- 870
Db	1932	--MNDSGLNTAIKDLIKQKGLFTGLVGLGAGLGLQIDLQNPKEKIGMSINDLSKXGLF 1989
Qy	871	NTVLQWIDVASALHTMPSVIRTLNIRYVVTALNKAESNLPSWDEWQTLAENMEAGLSTQQ 930
Db	1990	NOITGFISANDIGQVISWMLQDIK-...PS-...NALKNDVAALGKQM 2030
Qy	931	-AQTADYTABRLSSVLNCFWFLANIQEGVSLHSDRLYSYFLIDNQVSSAIKTRLAEA 989
Db	2031	IGFELQDQTLNLSLESLQNOQIKSVLDKVAAGKLGPIYEQGLGD-...LIPN 2079

Qy	990	IAGIQLYINRALNRI-...EPNARADVSTROFTDWTNNRYSTWGGVSRVLYVYENY 1043
Db	2080	LGKGLFPYGLSQVMQKGFDFSNAGQNVFV-QNSTFSNANGGTLFSNAGNSLIFAGNNH 2138
Qy	1044	IDPTQRIQTRMDELLENISQSKLSRDYVEDAFKTYLTTFETVADILKVVSAYHDNUNSN 1103
Db	2139	IAFTNHAGTLQLLSQDVSNIN-...ITTLNASNGLUK-INAAANNVSVS 2181
Qy	1104	TGLTWVF-...GQTRNLPYVWRNVDISRMOA- 1132
Db	2182	QG-NLFSVASCQAQSDPTTANIANPCALSAQSTNGASSNANAPALSNNDESMLVAA 2240
Qy	1133	-----GELAANAWEKWKIDTAVNPYKDAIRPVI FRERLHLIWEKEEVAKNGTDPVET 1186
Db	2241	NDFNFSGNIYANGVDFSKIKGSAN-... 2265
Qy	1187	YDRPTLKLAFIRHDSWSAPWSYDITTOVEAVTDKPD--TERLALAASQFQGBDITLLVP 1244
Db	2266	-----IKNLYLYNNAQFQA--NNLTISNOAVLEKFNASFTVNNLNIQGA-FNNNATQKIE 2316
Qy	1245	VYK-...TGKSYSDFGSNKNVAGMTIYDGSFKKMENTAL-SRYSOLKNTFD 1292
Db	2317	VLQNLVIASNASTGIYGLEVGALNNSGAIHF-...NLENTQTPPLIOAEGIIN 2369
Qy	1293	IIHTQ-...GNDLVRKASYRFAQDFE-VPASINMGSAIGDDSLTYMENGNIPOITSK- 1344
Db	2370	LNTTQTFPMVNNSMANNNTYTLKSSRYIDYINPNLSQSYLNLYTLININGNHIIEKN 2429
Qy	1345	YSSDNIAITLHNAFTVRYDGGNVRNKOIS-...AMKLT- 1381
Db	2430	GALTYLQORVLLQDKGLLSVALPNSNNSQNNILSLSVLYNQVMSCGDKAMDFTPTTL 2489
Qy	1382	---GVDGKSOY-...GNFIANTYKHYGSDLGCPITVYNTKNTNYIASVQGH 1429
Db	2490	QDYIVIGIQGSALNQIEAVCGNALKWLSTLMETKENPPFAPILYKXHSLEILGVTKDL 2549
Qy	1430	MN-...ADYTRRLILTPVNNYIARLFEFFPSPNTIILNTVFTVG 1469
Db	2550	QNTASLISNPNFRDNATNLLELASYTQTSRLTKLSDFRSREGESDFSLELKNKRFSDP 2609
Qy	1470	SN-...KTSDFPK-...CSYAVDGN-...SQFQ-...IFSYQSSGW 1502
Db	2610	NPEVFKVYSQLSKHPNMLWQVGGASFISGNGTLYGLNAGYDRLVKNVILGYSVAYG 2669
Qy	1503	LDIDTGI-...NNTDILKITVMAGSKTHFTASDHIAASLPANSFDPMPYTFKPLEIDASS 1557
Db	2670	SDFNGNIMHSLGNVVDVGMVARAPLKRNEFT-...LSAN-... 2704
Qy	1558	LAFTNNIAPLDIVFETKADGRVLGKI KQTLISV--KRVNYP-...EDILFLRE 1605
Db	2705	-----ETYGGNATSSINSSNLSLVLNQRYNYNTWTSVNGVGYDFMFKQK 2750
Qy	1606	T-...HSGAQYQWOLGVYRIRLNLTLASQLVSRANTCIDILTMETORLPEPPLGEG--F 1658
Db	2751	SVLKPQVGLSYHFGLGSKMGKNDAAKQFLMHSNPSNESVLTNLMGLESRKRYFGKNSY 2810
Qy	1659	PANFVLPKYPDAEHDGRWFKIHNVGNGTGPYYSGLMSDTSMTLFFVYASGY 1718
Db	2811	FVTARLGR-...DLIKSKGSNTVPFVGENTLILYRKGEVNT--FASVITGGE 2857
Qy	1719	MH-...EGVRLGVGYQKI 1732
Db	2858	MELWRLVYVYVAGVGLKXGLQYQDI 2881

RESULT 4
 S55805
 alpha-toxin - Clostridium novyi (ATCC 19402)
 C/Species: Clostridium novyi
 A/Variety: ATCC 19402
 C/Date: 28-Oct-1996 #sequence_revision 08-Nov-1996 #text change 09-Jul-2004
 C/Accession: S55805; S71294; S71158; S44273; I40834; S44272
 R/Hofmann, F.; Herrmann, A.; Habermann, E.; von Eichel-Streiber, C.

Db 1666 KDLNLYIDSVIIIPKVVTKKIIHPIPNPQIMINTQSIHDKCHLI-----IDSVLTN 1720
Qy 1540 SFDAMPYTFKPLETDASSLATNIAPILDIVPETHAKDGRVLGKIKQTLVKRVNYP-- 1597
Db 1721 NYH-----WESDGDLLIITNG---LDI--NIRILQGLSFGPKYKNVILKFSNYDELS 1767
Qy 1598 -EDILFLRETHSGAQYM-----QLGVYRIRLNTLLASOLVSRAANTGIDTIL 1642
Db 1768 LNDPELLQNYNVKGLYINGELHYKNIPODTEYEWINDSRWYFDSINLIKAGYQEI- 1826
Qy 1643 TMTQRLPEPPLGRGFAN-----FVLP---KYDPAEH-GDERWPKIHIGNVGN 1688
Db 1827 -----EGERYFNPNTGVQESGVFLTPNGLEYFTNKHASSKRW----- 1864
Qy 1689 TGROPYSGMLS-----DTSFSTMTLFPVPAEG-YMHGCVRLGVGYQ-----KITV 1734
Db 1865 -GRAINTGWJTLDNKYYFOSNSKAVTGLQKISDKYYFNDQGMQIKWQIINNKKYF 1923
Qy 1735 D-NTWESAFFYFDETKQFV-----LINDA-----DHDGQMTQGI 1769
Db 1924 DQNTGEAIIHFNKKERYFDSGRLLTGYOVIGDKSYYPSDNINGWEGSGVLKSGI 1983
Qy 1770 VKNIKKYKGLNV--SIATGVSAPMDFNSALYYWE-----LFYYPMM 1812
Db 1984 FKTPSGFKLFSSEGDKGAINYKGLWDLGNK--YFNSDSIAVTGYSYNIKGIQYFNPKT 2041
Qy 1813 CFQRLLEKQF--DEATQWYNYVNPAGY 1839
Db 2042 A---VLTNGMYTLDNNYYVNSGHNVLGY 2067

RESULT 5
I40884
cytotoxin L - Clostridium sordellii
C;Species: Clostridium sordellii
C;Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 09-Jul-2004
A;Reference number: I40884; MUID:95369733; PMID:7642137
A;Accession: I40884
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2364 <RES>
A;Cross-references: UNIPROT:Q46342; UNIPARC:UPI00000B02BE; EMBL:X82638; NID:g1000694; PI
C;Keywords: cytotoxin

Query Match 1.8%; Score 229.5; DB 2; Length 2364;
Beat Local Similarity 18.5%; Pred. No. 0.0018;
Matches 429; Conservative 308; Mismatches 711; Indels 873; Gaps 119;
Qy 23 ADLOYLSFSELKTFDDQLSNGEARHLYHETIEOKNNRLLERARIFTRANPQLSGAIRLG 82
Db 7 AQLQKVVVFPRIQEDYVAINALALEYHNMSSESVVEKYLK------ 50
Qy 83 IERSVRSYDEMGARSSSPKGVASMPSPAGYLT-----ELYREA-----KDIHF-- 131
Db 51 -DINNLTNDY-----LNTYKSGRNKALKKPKFXYLTMEVLELKNNSLTPEVKNLHFIW 102
Qy 132 -----SSAYHLNRRPDLADLTLSQGNMDPEISTLTLSNELLLEHITRKTGGSDALME 186
Db 103 IGGQINDTAINYINQKDV-----NSDYTVKVFVDSNAFLNTL-KKTIVES-ATNN 152
Qy 187 SLSTYQAIDTP-----YHQPVEFIRQVI---MTH---DSTLSALSRNPVEMQAGASLLAI 238
Db 153 TLESFRENLDPEFDYKNFYKRMELIYDKQHEIDYKYSQIENPFP-----I 201
Qy 239 LAN-ISPELYNLITEI-----TEKNADALFAQNFSENITPENFASQMIATKYVLEL- 290
Db 202 IDNIKTYSNEYSKDLALNKYIEESLNKINTANGNDIRNLEKFADED-LVRVLYNQELV 260
Qy 291 -----SEVQKYLGLMQLONGYSYSTAYVD-NISTGL-----VVNSES----- 325

Db 261 ERWNLAASDILRISMLKEDG-----GVYLDVDILPGIQDLPFKSINKPDSITWTSWEMI 315
Qy 326 KLEAYKTRVKTDDY-----DKNINYFDLMVEGNNQFFIRANFKVSRFEGATLAKNAGPS 380
Db 316 KLEAI---MKYKEIYFGYTSKN---FMDLDE-----EVQSFESALSCKSDKS 357
Qy 381 GIVGSLG---GPL-----TANT-----NPKSNYLSNISDSEYKNGVKIYAYRYTSS 423
Db 358 EIFLPLDDIKVSPLEVKIAFANNNSVINQALISLSDYCSDLVINQIKRKYKILNDNLNPS 417
Qy 424 TSATNOGGGIFTFESYPLTIFALKI-----NKAIRLCLTS-----GLSPNELOTVIRS- 471
Db 418 ---INEGTDFNT---TMKIFSDKLASI SNEDNMFMKIKITNLYLKVGFADPVRSTINLSG 470
Qy 472 -----DNAQGI-----INDSVLTKVFFYTLFVSHRY----- 496
Db 471 PGVYTGAYQDLLMPKDNSTNIHLEPELRNFPFPKTKISQLTQEIEITSLMSFNQARAKSQ 530
Qy 497 -----ALSFFD---AQVLNGSVINQVADDDSVSHFNRLFNTPPLKKG----- 535
Db 531 FEEYKKGYPGALGEDDNLDFAQ-----NTVLDDKYVS--KKILSMKTRNKEYIHYI 581
Qy 536 -----IFEADGNTVSDIDPEE---QSTFARSALMRGLGVNGSGELYQLGKLAGVILDAQ 584
Db 582 VOLQGDKI SYEASCNLFSDPYSSILYQKNIEGSETAYVYVVADEAIKEIDKYRIPYQIS 641
Qy 585 NNTILSVFVSSLYRLTLARVHQLTVNELCMLYGLSPFNKGKTTASISSEGLPLRLVILWY 644
Db 642 NKRNI-----KLTFIG-----HGKSEFNTDTPFANLDVDSLSSEIETIL 679
Qy 645 QVTQWLTEAREITTE-AIWLTLCTPEPSGNISPE-----ISNLNLNLRPISBDM 692
Db 680 N-----LAKDISPKYIIEINLLGCMWFSYSISABETYPCKLLKIKDRVSELMPSISODSI 735
Qy 693 QSHNRLEQABILAPPTIAATLHLASPDMARYIL-----LWTDNLRPGGLDIAGFMTLVKE- 747
Db 736 TVSANQVEVRI-----NEEGREILDHSGKWINKESIIKDSS-----KEY 777
Qy 748 -SLNANETTQIV-QFCHVMAQLSVQTLR-----LSEAELSVL----- 784
Db 778 ISFPNKENKIIVKSKYLH---ELSTLQEIERNNANSSDIDLEKKVMLTECEINVASNDR 834
Qy 785 -VLSGFAVLGAKQOPAGQHNI-----DTLPSLYRFHOWINGL----- 820
Db 835 QIVEG-RIEEAKNLTSDSIYNIKNFKLIBSISLSYDLK--HQ--NGLDSSHFI SFEDI 889
Qy 821 -----GNP-----GSDTL-----DMLRQQT 835
Db 890 SKTENGPRIRPINKETGNSIFIETEKEIFSEYATHISKEISNIKDTTFDNNVNGKLVRKVN 949
Qy 836 LTA-----DRLASVMGLDISWVTOA---MVSAGVNOLOCMQDINT 872
Db 950 LDKAAHEVNTLNSAFFIQLSLEYNTTKESLNSLVAMKVQVYVQALFSTGLNTI---TASK 1006
Qy 873 VLOWIDVA---SALHTMPS-----VIRTLVN-IRYVTALNK-ABSNNLPSMDEWOTLAENM 922
Db 1007 VVELVSTALDETIDLLPTLSEGLPIIATIDGVSGLGAIKELSETNP-----LLRQEI 1060
Qy 923 EAGLSTOQAQTLADYTAERLSSV-LCNWFILANIOP-EGVSLHSRDDLYSYFLINDQVSSA 980
Db 1061 EAKIGIMAVNLTAASTAIVTSALGIASGFSILLVPLAGISAGIPSLVNNELILODKATKV 1120
Qy 981 I---KTRRLAEATAGIOLYINR-----ALNRIEPNA-----RADVSTRQFF 1018
Db 1121 IDYFKHISLAEAGFTLLDDKIIMPDDLVLSIEDFNNNSITLKGKEIMWAEAGSGHTL 1180
Qy 1019 TDMTVNNRYST-----WGGVSRVLYYPENYID-----PTQIRIGQT----- 1053
Db 1181 TD-DIDHFFSPSPITYKPKWLSIYDVLNIKKEIDFSKDLMLVLPNAPNRVFGEMGWTPG 1239
Qy 1054 -RMMD-----ELLENISQSKLSRDTVEDAFKTYLTRFTETVAD---LKVVSAYHD-NVNSNT 1104

Db 1240 FRSLDNGTKLLDRI-----RDHYEQGF--YWRFAFIADALITKLPKRYEDTNRINL 1291
 Qy 1105 GLTWFGVQTRNLPEIYWRNVNIDSRMQAGELAAWAKWTKIDTAVNPKDAIPVIFRE 1164
 Db 1292 D-----GNTR-----SFIVPVI--- 1303
 Qy 1165 RLHLIWEKEVAKNGTDPVYDRTFKLAFRLHDSGWSA-----PW 1207
 Db 1304 -----TTEQIRKN-----LSYFYSGSGSYSLSPYNNMIDNLVENDTW 1344
 Qy 1208 SYDITTOVEAVT---DKKPDTERLALASGQGEDTLLVFFVYKTKSYSDSGSKNKNVAG 1264
 Db 1345 VIDVNVKNTIIESDRIQKGLIENILSKNIEDNKILNNHTNFYDGINESNRPTIS- 1403
 Qy 1265 MT-----IYDGSFKKMENTALSRYSLKNTFDII----- 1294
 Db 1404 LTFPSEIDINIIIEDLVSKYKILLSGNCMKLIENS-----SDIQKIDHGFNGEHQK 1458
 Qy 1295 -----HTQGNLV--RKASVRFQDPE-----VPASLNMGSAIGDD--SLTV 1332
 Db 1459 YIPYSIDNETKYNGFIDYSKKEGLFTAEFSGNESIIRNIYMPDSNNLFIYSSKDKDIRI 1518
 Qy 1333 MENGNIPOITSKYSSDNLAIITLHNAAFVRYDGSNGVIRNKQISAMKLTGVDGSKQYQNA 1392
 Db 1519 INRGDVKLLIGNYFKDKDKVSL---SFTI-----EDTNTIKLNGV---YLDE 1559
 Qy 1393 FIANTVKHYGGYSDLGGPITVYAKTKNYIASVOCHLMNADYTRRLIITPVENNYARLF 1452
 Db 1560 NGVAQILKF---MNAKSAINTSLSLMMFLSIN-----IKNIFYNNL- 1599
 Qy 1453 EFPSPNT--ILNTVTV-GNKTSDF-----KKCSYAVDGNNSQG 1490
 Db 1600 ---DNIEFTLDNFIISGNSISQGFELICDKDKNIQPYEINFKIKETSYTLVGNRQN 1655
 Qy 1491 FOIFESSQSSWLDIDGTINNTDKITVMAGSKHTFTASDHASLPANSDAMPYTFKP 1550
 Db 1656 LIVEPSYH---LD-DSG---NISSTVINFSQKYLIGIDRYV---NKVIIAPNLYTD 1701
 Qy 1551 LEIDASSLAFTNNTAPLDIVETRAKQGRVLGKIKQTLKVRNVNVEDILFLRETHSGA 1610
 Db 1702 -EINITPYKKNYICPEVIILDANYINEKI-----NVINDLSIRYVD----- 1744
 Qy 1611 QYMQGVYRIRLNTLLASQLVSRANTGIDTILTWETQRLPEPPLGEGEFANFVLPKYDPA 1670
 Db 1745 -----NDGSDILILANSERDNPQV----- 1764
 Qy 1671 EHGDERWFKIHGNGVGTGQPYSGMLSDTSETMTLV-----PYAEGYMYHE--- 1721
 Db 1765 -----KIRFVNFKSDTAADKLSFNFSQKQDVSVSKIISTFSLAAYSDFDYBFLG 1816
 Qy 1722 -----GVRLGVGY-----QKITDONTWESAFFYDETKQQPVL 1754
 Db 1817 VSLDNDFYINSFGNMGVGLIYINDSLYFKEPKNNLITGTTIDGNKYFYDPTKSGNAS 1876
 Qy 1755 INDADHDGSG---MTQQQGVKKNYKYGFLNVSIAIGYSAPM-----DFN 1795
 Db 1877 IGEITIDGKYFNKQGILO-----VGVIINTSDGLKYPAPAGTLDENLEGESVNFPIGLN 1931
 Qy 1796 SASALYYWELFYTPMCMQFRLLOEKQFDEATQMINVYNP 1836
 Db 1932 IDGKIYYPEDNY-----RAAEVWKLDDDET-----YFNP 1961

RESULT 6
 D64962
 probable membrane protein b1978 - Escherichia coli (strain K-12)
 C:Species: Escherichia coli
 C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
 C:Accession: D64962
 R:Blattner, P.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
 .A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617; PMID:9278503
 A:Accession: D64962
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-2383 <BLAT>
 A:Cross-references: UNIPARC:UPI0000047C86; GB:AE000289; GB:U00096; NID:gl788285; PIDN:A
 C:Keywords: nucleotide binding; P-loop; transmembrane protein
 F:54-70/Domain: transmembrane #status predicted <TM>
 F:1564-1571/Region: nucleotide-binding motif A (P-loop)

Query Match 1.8%; Score 229.5; DB 2; Length 2383;
 Best Local Similarity 18.1%; Pred. No. 0.0019;
 Matches 491; Conservative 352; Mismatches 880; Indels 983; Gaps 127;

Qy 208 QVIMTHDSITLSALSRNEVVGQAGSLLATLANISPELYNLTETETKKNADALFAQNF 267
 Db 85 QIATANTV-----PYTLGALESAQSA-----ERFGISVAELRKLNQPRFARSF 131
 Qy 268 -----SENTIPENFASQSWIAKYKYLESEVQKYLQNGYSDSTSA- 310
 Db 132 DNVKQGDDELDPVAPQVSEKLTTPPGNSNDLQ-----QIASTSQIGSLLAEDMNSQAA 187
 Qy 311 -----YVDNISTGLVANNESKLEAYKTRVKTDDYDKNINYFDLM---YEGNQPFIRAN 362
 Db 188 NMARGWASSQASGAMTDWLSRFGTARITLGVDEDFSLKNSQDFLHPWYETPDNLFFSQH 247
 Qy 363 FKVSREFGATLRKNAQSGIVGSLGGLIANTNFKSNLYNISDSYKNGVKIYAYRYTS 422
 Db 248 -----TLHRDERTQINNLGW-----RHPTPTWMSGIN-----PFDDHDL 283
 Qy 423 STSATNGGGIIFTPESYPLTIFALKNAIKARLCUTSLGSPNELQTIIVRSDNAQG--IIND 480
 Db 284 SRVHSRAGIGAEYWRDY-----LKLSSNGYLRLTNRSAPELNDYEAPANGMDVRAE 337
 Qy 481 SVLT-----KVEYTLFYSHRYALSFDDAQVLSGVIN---QYADDDSVSHENLENTP 530
 Db 338 SWLPANPHLGKLYVEQYGYDEVALFDKDRQSNPHAITAGLANTYTPPLMT-----FSAE 392
 Qy 531 PLKGIKFEAD-----GNTV--SIDPDEBQSTFARSALMRGLGVNSGELYQLGKL 577
 Db 393 QRQKQKQENDTRFAVDFTWQPGSAMQQLDNE---VAARSL-----AGSRVYL--- 439
 Qy 578 AGVLDAQNTITLSVFVYSSLYRLTLARVHQLTVNELCMLYGLSPFNKTKTASLSGELP 637
 Db 440 ---VDNNNIVLE-YRKKELVRLTL-----TDPVTGK-----SGSVK 472
 Qy 638 RLVTWL-----YQVQWLTAE-----ITTEAIWLLCTPEFSGNISPEISNLLNLRP 685
 Db 473 SLVSSLOTKYALKGYNVEATALEAAGKVVTGKDILVTLPAYRFTSTPETDN----- 525
 Qy 686 SISEDMAQSHNRELAELAPFIAATLHLASPDMAVYILLWTDNLRPGGLDIAGFMTLVL 745
 Db 526 -----TWPIETAE-----DVKG----- 538
 Qy 746 KESLNANETTLQVQFCHVMAQ-----LSLSVQTLRLSEALSVLVIGFAVLGAKNQPAQ 801
 Db 539 --NLSNREQSMVVQAPTLTSQKSSVLSITQLN-----ADS 573
 Qy 802 HNIDTFLSLRFHWINGLGNPGSDTLDMLRQQLTLADRLASVMGLDLSMTVQAMVSAGV 861
 Db 574 HSTATLTFFIAH-----DAAGNP-----VVGLVLSTRHE-----GV 603
 Qy 862 NQ--LQWQD-----INTVLQWIDVASALHTWPSVI-----R 891
 Db 604 QDITLSDWKNGDSGYTQILTTGAMSGTLTLPQNGVDAKA---PAVNVILSVSSSR 659
 Qy 892 TLVNI-----RY-----VTALNAKESNLPSPDEWQTL-----AENMEAGLSQQAOTL 934
 Db 660 THSSIKIDKRYLSGNPIEVTVELRDENDKPKQKQQLNNAVSDINVKPGVTTDWKETA 719
 Qy 935 ---ADYTAERLSS-----VLCNWFLANTQPGVSLHSRDDLYSY-FLTDNQVSSAIK 982

[illegible]

Db 1886 NN--DGVMLGVFKGPDG-FEYFAPANTQNNIEGOAIVYQSKFLTLNGKKY----YFDN- 1938
Qy 1101 NSNTGLTWVGTQRENLEPEYWR-----NVDLSRMOAGELANAW 1140
Db 1939 NSKAVTGWRI---INNEKYYFNPNNATAAAGLVQVIDNNKYYFNPDTAIIISKGWQTNGS 1994
Qy 1141 KEWKIDTAV--NPVKDAIRPVI FRERLHLTWEEKAVK-----NGTD---PVETD 1188
Db 1995 RYFEDTDTAIAFNGTKTI-----DGKH-FYFSDDCVVKIGVPSFSTNGEYFAPANTYN 2046
Qy 1189 R-----FTLKLAFLRHDSGSAFWSYDITQVEAVT-----DKKP--DTERLALAASGF 1235
Db 2047 NNIEGOAIVYQSKFLTLNGK-----KYYFDNNSKAVTGWQIDSKKYYFNTNTAEAAATGW 2101
Qy 1236 QGEDTLVVFYVTKGSYSDFGSGSNKVAGMTIYGDSFKGMENTALSR--YSOLKNFTDI 1293
Db 2102 QTID-----GKKYYFNTTAEAAATGWQIDGKKYYFNTTALIASTGVTIINGKHFY 2152
Qy 1294 IHTGNDLVRKASVRFADFEVPASLNGMSAIGDSDLTMENGNIPQITSKY--SSDNLA 1351
Db 2153 FNTDG--IMQIGVFKGPNGFYFAPANTDANNIEGOAILYQNEFLTLNGKKYYFGSGSKA 2210
Qy 1352 ITLHNAFTVRYDGSNNVRNKQ-----ISAMKLTGVDGKSQYGNAFIANTVRHYG 1403
Db 2211 VT-----GWRILNKKYYFNPNNALAIHLCTI-----NNDKYYF 2245
Qy 1404 GYSLGGPITVYNTKKNYASVOGHLMNADYTRRLILTPVNNY-----ARLFE 1453
Db 2246 SY-----DGTILONG-----YITIERNNFYFDANNESKMYTVGVFK 2279
Qy 1454 FP-----PSP-NTILNTV-----FTVGSNKTSDFKCSAVDQ-----NNSQGFQ 1492
Db 2280 GPNGFYFAPANTHNNIEGOAIVYQNKFLNGKKYYFNDSKAVTGWQIDGKKYYFN 2339
Qy 1493 IFSSYQSSGWLDDID-----TGINTDIIKITVMAGSK-----THPTASDHIA 1534
Db 2340 LNTAEAAATGWQIDGKKYYFNLTAEAAATGWQID-----GKKYYFNTNTFIATGYT 2392
Qy 1535 SLPANSP-----DAMPYTFKPLEIDASSLAFTNNIAPLDIVFETK----- 1574
Db 2393 SINGKHFYFNTDGMQIGVFKGPNGFY-FAPANTDA-----NNIEGOAILYQNKFLTL 2445
Qy 1575 -----AKDGRVLGKIKOTLSVKRVNYPEDILFLRETHSGAOMQLGVYRIRLTLA 1627
Db 2446 NGKKYYFGSDSKAVTGLR-TIDGKKYYFNTTAVAV-----TGWQINGKKYYFNTNSIA 2500
Qy 1628 S---QLVSRANTGIDTILTMETQRLPEPLGEGFPANFVLPKYDPAEHG-----ERW 1677
Db 2501 STGVTIISGKHFFYFNTDGMQIGVFKGP---DGF--EYFAPANTDANNIEGOAIRYQNR 2555
Qy 1678 FKIHIGNVGGNTGROPYYSGLMDSMTSETMLFVPIYARG--YYMHEGVRLGV-GYQKIT 1734
Db 2556 LYLH-DNT-----YFNGNSKAATGWVTI-----DGNRYFPEPNTAMGANGYKTI 2601
Qy 1735 DNTWESAFYFDETKQFVLINDADHSGMTQOQIGVKIKKYYKGLNVSATGYAPMDF 1794
Db 2602 KN-----FYF-----RNLQIQIGVFKGNSGFE-----YFAPANT 2630
Qy 1795 -----NSALYYWE-----LPYTPMCMCFORLLOSKQ 1822
Db 2631 DANNIEGOAIRYQNRFLHLGKIYYFGNNSKAVTGWQINGKYYFME----- 2678
Qy 1823 FDEATQWINYVNPAGYI-----VNGEIAPIW 1850
Db 2679 -DTAMAAAGLFEIDGVYFFPGVDGVKAPGIY 2709

RESULT 8

F90696
hypochemical protein ECs0542 [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: F90696

R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90696
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-5291 <HAY>
A:Cross-references: UNIPROT:O8X2T1; UNIPARC:UPI0000011021F; GB:BA000007; PIDN:BA033965.1;
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0542

Query Match 1.7%; Score 222; DB 2; Length 5291;
Best Local Similarity 18.8%; Pred. No. 0.02; 988; Indels 956; Gaps 135;
Matches 532; Conservative 353; Mismatches 988; Indels 956; Gaps 135;
Qy 146 ADLTLQSNNMDTEISTLTLSNELLEHITRTKGTGSDSALMESLSTYRQAIDTPYHQPYET 205
Db 392 ADGWSVSLPTADLSNLTASQYTVSASVSDKAGNPASA-----NHGLAVDLTV--PVL 443
Qy 206 IRQVIMTHDSTLSALSRRNPEVMGOAE-GASLLAILANISPELYNLTETETEKADALFA 264
Db 444 INTV-----SGDDIIINAAEHGOALVIGSSSTGSGEAGDVITVTLNSKTYTTLMD 491
Qy 265 QNPSENITPNFASOSIAKYKYLESEVQKYLGMQLONGYSDSTSAYV-----DNISTG 318
Db 492 -----ASGNNSV--GVPAADVT-----ALGSGPQITTAATDAAGNSDDASRT 532
Qy 319 LVVNESKLEAYKITRVTKTDYDKNINYFDLMYEGNNQFFIRANFKVSREFGATLR---K 375
Db 533 VTVNLAA--PIIGINTIATDDVIK-----ATEKGADLQITGT 567
Qy 376 NAGPSG--IVGSLGP-LIANTNPKSVNLSISSEYKNGVKYIYARYTSTSTATNOGG 432
Db 568 SNQPAQTITITVLNGQNYTATDTSNGNWSATVPASAVS---ALGEANYTVTANVTDTAGN 624
Qy 433 IFTESYPLTITFALKLNKARCLTSLGSLPNELOTVRSDDNAOGII-----NDSVLT 484
Db 625 SNSASHNVL-----VNSALPAVTINAVATDIIINAAESGNAQITISQVTAAGQDVTTV 678
Qy 485 KV-----FYTLFYSHRYALSFD---DAQVL-----NGSVIN----- 512
Db 679 TLGGNTYATVQSNLSNSVDVPAADIQALNGDLTVNASVTNGVNTGSGSRDITIDANL 738
Qy 513 -----QVADD-----SVSHFNRLFTNPPKGIKEADGNTVSIIDPEEQSTFARSALMRG 563
Db 739 PGLRVDTVAGDDVINSIEHNQALVITGSSG-LTAGTALTVEIN---NVTYGAIVLADG 793
Qy 564 LGVNSGELYQLKLAGVLDAQN---TITLSVFVISLRLTLARVHQLTVNELCMLYG 619
Db 794 -----TWSLGVPA--VDVSNWPAGTVNIIVSGTNSAGTSTI--THPVTVDLAGVAIT 842
Qy 620 LSPENGK-----TTASLSSGELPRLVIMLYQVTLQWLTETABITTEAIV----- 661
Db 843 INTLSGDDVINAVEKGETLVVSGTSGVEAGQTVTVTFGGKNYTTVE--ANGSWTNVP 900
Qy 662 ---LLCTPESGNTSPEISNLLNNLRPSISDMSQAQSHNRELOAEILAPPIAATLHLASPD 718
Db 901 PADLAALPDGAGNVQASVSNINGN---SAQADRAYS-----VDATAFLVTINT-IASDD 950
Qy 719 MARYILLMTDNLNRPGLDIDAGFMTLVLKESLNA---NETTQLVQFCHVMAQLSLSVQTLR 775
Db 951 ILNV-----SEAGAGTISGTTTAQAGQTLTVTLNNNTYQTT---VLADGTWSV---N 997
Qy 776 LSEAEISVLVISGFAVLGAKNQAPQOHNIDTLFSLYRPHOWINGNLPGSOTDLMRLROOT 835
Db 998 VPAADLSGLTASSVTVTATVSDKA-----GNPASADHALVVDIT 1036
Qy 836 LTAERLASVGLDIDLSMTV---QAMVSAGVNOQCQWQDINTVLOWIDVASALHTWPSVIRT 892
Db 1037 APDLTINTVAGDDIINAIEHGQALVSVGSTGAAAGDVVVTV-----T 1078

Db 1483 -VNYLLDDIV---KIHDSFFL-----RIQLEQ-LSHVLVDKITAVLPKSLTIK 1526

RESULT 13

S59413

probable membrane protein YLR454w - yeast (Saccharomyces cerevisiae)

N;Alternate names: hypothetical protein L9122.5

C;Species: Saccharomyces cerevisiae

C;Date: 30-Nov-1995 #sequence_revision 23-Feb-1996 #text_change 09-Jul-2004

C;Accession: S59413

R;Kirsten, J.

submitted to the EMBL Data Library, March 1995

A;Description: The sequence of S. cerevisiae cosmid 9122.

A;Reference number: S59414

A;Accession: S59413

A;Experimental source: strain S288C (AB972)

C;Genetics:

A;Gene: MIPS:YLR454w

A;Cross-references: SGD:S0004446

A;Map position: 128

C;Keywords: transmembrane protein

F;7-23/Domain: transmembrane #status predicted <TM1>

F;173-189/Domain: transmembrane #status predicted <TM2>

F;835-851/Domain: transmembrane #status predicted <TM3>

F;2569-2585/Domain: transmembrane #status predicted <TM4>

Query Match 1.7%; Score 217; DB 2; Length 2628;

Beat Local Similarity 17.9%; Pred. NO. 0.011;

Matches 511; Conservative 398; Mismatches 1000; Indels 940; Gaps 136;

Qy 144 DLADLTLSQSNMDTEISLTLSNELLEHTRKGTGSDSDALMESLSTYRQAIPTYHQP- 202

Db 215 NLLNLFINKENVND-----LMSNEKLQRLADTTKANBELKDEDIAGKQ--DLVYAMEK 266

Qy 203 ---YETIRQVIMTHDSTL---SALSRLPEYMGQAGASLAILAILANISPELYNLTIERIT 255

Db 267 FVDRIKPLKENNVTVKLIQKDFLTHNPHELLGMNKYISYNVLVSNL-----NFTNRF- 320

Qy 256 EKNADALFAQNPSENITPENFASQSWIAKYGYGLESEVQKYLGLMQLNGYSDSTSAVDNI 315

Db 321 -RNEMPGYTLIFEERDSPEKFS--IIMARF-----NI 349

Qy 316 STGLVNNESKLEAVKITRVKTDYDKNINYFDLMYEGNQFFIRANFKVSRFEGATLRK 375

Db 350 YLNLNRKHQSHAKQKLTIEI-----PNVSIF-----GETNLFQO-KFRLSNLHAKELE 397

Qy 376 NAGPSGIVGSLSGPLI---ANTNPKSNLYSLNI---SDSRYKNGVKIYAYRTSTSATN 428

Db 398 NA-IFNIKGNISLTDMDPNVISIPKFLSNIKVFTSSCPKNKI---LKENSHVKFLT 452

Qy 429 QGGGIPTSPESPLTIPALKKNAIRLCLTSLGSLSPNELQTVIRSDNAQGIINDS---VLTK 485

Db 453 RRRVLFDFYKFCPLINMKFT-----LDDPKFVINDDKODLIIGK 491

Qy 486 --VFYTLFYSRYALSFPDDAQVLNGSVINQYADDSVSHENRLEFNTPLPKGKIPEADGNT 543

Db 492 FSVFMISHSKRYTL-----GNN 509

Qy 544 VSIDPDERQSTFARSALMRLGVNSGELYQLGKLAGVLDAQNTITLSVFVSIYRLTL 603

Db 510 LMEKEEFTQHIFYESHW-----NVELLDM-KLQHIKHQK-----YEHTIL 549

Qy 604 ARVHQLTVNELCMLYG--LSPNGK--TTASISGSGELPRLVITWLVOYQWLTEAEITTEAI 660

Db 550 -RVDISIAEEKVQLLPDILCSANADITLMLDLSLPTWMLSELVHLNLSQLANVEENY 608

Qy 661 WLLCTPESGNI---SPRISNLLNLRPSISGDMQAQSHNRELOAEIL--APFIAATLHLAS 716

Db 609 KFEYFEKASNLQNMKAECNSNAKCLR-----QKEILSPDFMFOOL----- 649

Qy 717 PDMARYILLMTDNLRPGGLDIAG-----PMTVLVKESLNANETTQLV-----QFCHVM 764

Db 650 PDFDYIKI---NIR-----DISSTLGARSVMPDRFSSVSDSQSKDLIDGKLRKYCNTV 702

Qy 765 AQLSLSVQTLRLSEALSVLVIGSFAVLGAKNQPAQHNIIDTLFSLYHFHFWINGLGNPG 824

Db 703 EKQI-----ALFGDKT-----QHNKI---G 721

Qy 825 SDTIDMLRQOQTLT-----ADRLASVMGLDISMVTQAMVSAGVNOLOQCOWDINTVLQ 875

Db 722 SNHATVRSGLTWFNSKDNKQPNHKSSIALDDISTSDA---TEVNLH--W-NINLLVN 775

Qy 876 WIDVASALHTWPSVIRTLVNIYRTALNKAESNLPWSDEWOTLAENMEAGLSTQQAOTLA 935

Db 776 --DITTSI-----IGTEPVESELS 793

Qy 936 DYTAEIRSSVLNCFNLANIQPEGVLSHRDLYSVFLIDN--QVSSAKTKRLAEAIAGI 993

Db 794 TKTVKVSNSIKLP-----PDTSFSSNESDSKILIQINHSGTSSVLSMISIFLAVSGI 848

Qy 994 QLYINRALNRIEPNARADVSTRQPTDWTNNRYS-----TWGGVSRLLVYYPENIDPTQ 1048

Db 849 HT-LNQIFGHCHQKMRQSKTKQYFLALSESKKCKSIKKG----- 890

Qy 1049 RIGQTRMDELLE-NISQSKLSRDTVEDAFKTYL--TRPETVADLKVVSAYHDVNSNTGL 1106

Db 891 -----QLKELLEINFSEYISQII--ALPNGLRTKPEPTSTP--ITVKNCTISVSGQ 939

Qy 1107 TWFGVQTRENLPEYVYR-----NVDISMQAGELAAANAKWKTKIDTAVNPYKD 1155

Db 940 YFRMVESPTQPNFWRMFCINGFKVMIHIDLLKQOMKKL--NSLQWKEKLSAITLEND 997

Qy 1156 AIR-----FVIF--RERHLIIVWEKEE--VAKNGTDPVETYDRFTLKL 1194

Db 998 SHWFSIPHPPEMFKIIDSIPFKSIQMLYSKATKDDLIIFPHKIETPLSLPKLK- 1056

Qy 1195 AFLRHGWSWAPWSYDITQVEAVTKKPDTERLALAAAGFGQGEDTLVVFYTKGKSVS- 1253

Db 1057 -----SKRWLF-----SISDDPLEALNTIFQIGLQERERLAKLQEFNKRLSE 1100

Qy 1254 DFGGSKNVAGMTIYGDGSPKKMENTALSRYSQKNTPTDIHTQGNIDLVRKASYRFAODF 1313

Db 1101 DLIKSQNAKEM-----KDDFEAIDNAI-----LKHRTGLWAKGKKELRKS----ATDS 1146

Qy 1314 EV---PASLNNGSAIGD-----DSL----- 1330

Db 1147 EIPLTALALINGKRDDRDRDRTQTFISPEIENAYNTLLANFSDSWIKRKYKVERREFD 1206

Qy 1331 -----TWMEGNIPQITSKYSSDNLAITLHN 1356

Db 1207 KNFSFLMGFTDYTKLPKDINKKVLFPFTNPFMLNLIENIDIDIIRPSCGIENIPNF1HD 1266

Qy 1357 AAFVTRVGDGNGVIRNQISAKMLTGVGDK-----SQYGNAPFIANTVKHYGYSDL 1408

Db 1267 V-----GKG-VPAKTEYSIMI PMHLDAKSEVVRWHLRDYPLFVSPPLSSTOSKETI 1318

Qy 1409 GGPITVYKTKNYIASVQGHLMNADYTRRLILTP-----VEN--NYVARLPEPPSPNT 1460

Db 1319 --PMRIYGD-----FMITEDMLQSDRELTLFVPLPSVTVENTDRYVSL-----FVPR 1366

Qy 1461 ILNT-VFT-----VGSNKTS-----DPKCSYAVDGNNSQGFQIF 1494

Db 1367 MTSAKIFTDLNFEINSNHTTRVTWGGSYQPAIQOTMQCLDNFSK--PPLDPSVLKGFWDK 1424

Qy 1495 SSYQSSGWLDI-----DTGINNTDITKITWAGSKTHTFTAS-DHIASLPSNPDAMPVT 1547

Db 1425 TRYLFHKGKINIVWKKRGKFEISLKGAKSPYMGGESAGFIVGFGNVNKLKNE--DNDPK 1483

Qy 1548 FKPLEIDASSLAFTNNIAPLDIVFETKAKGRVLGK-----IKQTLVSXRV 1593

Db 1484 FLSCSADKVFHSIPNYFAKPLLVMSRPSNTMTFIPNQODDTNMQRYASFYLLNTTSSKNE 1543

Qy 1594 NYPNEDI--LFLRETH-----SGAQYMQLVGVYRIRLNTLLASQLVSRANTGIDTILT-METQ 1647

Qy 768 SLVQTLRLSBAELSVLVISGFAVLGAKNQPAGQHNID---TLFSLYRPHQWINGLGNP 823
 Db 1094 SNNI-TINLDD---SVLSASNTSSLANINFGASQADFGGNTIIDTASFN-----1140
 Qy 824 GSDTLDMRLQOQLTRADRLASVWGLDLSMTQAMVSGVQLOQWODINTVLOWIDVASAL 883
 Db 1141 -FDSASSLNFNLTANGALNFNGYTPSLTKALMSVSGVFLGNGNDINL---SDINIFD 1195
 Qy 884 HTMPSVIRTLVNI-RYVTLAKA-----ESNLPSW-----DEWQ 916
 Db 1196 NITSVTVNINLAQGITGIGANGYEKILFVGMKIQNAVTSDDNNIOTWSFINPLNSQ 1255
 Qy 917 TLAENMEAG---LSTQQAQTLADYTAERLSSVLCNWFANLQPEGVSLHSRDDLYSYPLI 973
 Db 1256 IIQESIKNGDLTIEVLNPNASNTIPNIAPELYNQASKNPTG-----YSYDYS 1306
 Qy 974 DNQVSSAKITRLARAIAIGIOLYINRALNRIPENARADVSTQPFDTWTVNVRVSTWGGV 1033
 Db 1307 DNQAGTYLTSTNIKGLFTPKGSQTPQAPGTYSFPNQPLSSLIYNKGFSENKLTLLGIL 1366
 Qy 1034 SRLVYYPENIDPTQIGQTRMD--ELLENISQSKLSDRTVEDAFKYLTRFETVADLK 1091
 Db 1367 SONSATLKEMIESQDNITNINEVQLDDKIKITQVQKAL-----LETINHLT 1416
 Qy 1092 -----VVSAYHDVNSNTGLTFVVGQ-----TRENLPYYRNRVDIRMOAG 1133
 Db 1417 DNIQTFNNGNLIIGATQDNVTNSTSSIFGNGYSSPCLTDSATCSFRNTYL-----G 1471
 Qy 1134 ELAANAKWTKIDTAVNPYKDAIRPVIFRELHLIWEKEVAKNGDTPVETVDRFTLK 1193
 Db 1472 QLLG-----STSPYLGYNADPKAKSIYI-----TGT-----1498
 Qy 1194 LAFRLHDGWSAPWSYDITTOVEAVTDKDPTERLAALASFGQEDTLLVFYKTKSYS 1253
 Db 1499 ---IGSGNAFSGGSAUDTFO-----SANNLVNKANIEAQTUNIF-----1537
 Qy 1254 DFGSNKNVAGM---TIYDGSFKK-MENTALSRYSQKNTFDIITQGNDLVRKASYR 1308
 Db 1538 -----NLLQKGKIEKIFNQGSLANVLQVAMEKIKQAGLGNFENALSPLSK-----1585
 Qy 1309 FAQDFEVPASL---NMGSAIGDSDTVMEN-----GNI---PQITSK 1344
 Db 1586 -----ELPASLQNETLGLIQGNLDDLLNNSGVMAIQTIISSKLSIFGNFVTPSIEN 1640
 Qy 1345 Y-SSDNLAITLHNAFTVRYDGSNV-----IRNKQISAKMLTGVDGKSQYV 1390
 Db 1641 YLAKQSLKSLDDKGLNFIIGYMNABSLSSILSVLKDITNPPTSLOKDIGVANDLL- 1699
 Qy 1391 NAFTIANTVK---HYGGYSDL-----GGPITVYNTKNTYI---ASVQGHLMNADYTRRL 1438
 Db 1700 NEFLGQDVIKLESQGLVNIINNIIISQGLSGVYNGVGLSVLPPLSLQNALKENDL--GT 1757
 Qy 1439 ILTP-----VENNYARLFEPPFSP-----NTILNTVFTVGNKTSDFPKC 1479
 Db 1758 LLSRPLGLHDFWQKGVFNLSNGYFVWNSNSNATGSSILNFVANKSIIFNGDNTIDFSK- 1816
 Qy 1480 SYAVDGNNSQGFIPSS-----YQSSGWLDDTGINNTDI---KITV-WAGSKTH 1525
 Db 1817 -----YQALIPASNDVSNINITTILNATNGSLNAGLNNVSVQKGEICVNLANCPTT 1868
 Qy 1526 TFTAASHDIASLPANSFDMPTFFKPLEIDASSLAFTNNIAPLIDIVFETKAKDGRVLGKIK 1585
 Db 1869 KNSSTNSSVFTN-----ESLSVRANNFTFLGALASNGAIDLQSKNNSVI-----1915
 Qy 1586 QTLVSRVYNYPEDILFRETSHGSAQYQVLMQVYRIRLNTLLASQLVSRANTGIDTI---L 1642
 Db 1916 -----DTLNLNENAA-----LQANNLTITNAPNANSTANINGNF 1951
 Qy 1643 TWETO-RLPEPLGSGFPANFVLPKYDPAEHGDERWFKIH-----IGHVGN-----TGR 1691
 Db 1952 TINQOATLSTNAGSLNVNMGNF-----NSYGDVLVFNLSHVSVAIINAOQSATIMANN 2004

Qy 1692 QPYTSGMLSDTSETSMILFVYASGYMHGVRLOGVY-QKITVDNTWESAFFYFDETKQ 1750
 Db 2005 NPLIQNTSSKEVGTITLIDSAKIYY-----GYNNQITGSS-----LDNYLK 2048
 Qy 1751 QFVLIN-DADH---DSGWTQGGIVKNIKKYKFLNVSIATGYSAPMDFNSASALYYWEL 1805
 Db 2049 LYTLIDINGKMWTDNGLTNGQAVSVK-----DGLVVGFK-----DSQNYIYTSI 2097
 Qy 1806 FYTTPMCFQ---RLQOE---KQF-----DEATQWYNNVYNPAGYI 1840
 Db 2098 LYNKVKIANSNDPINNLQAPTLKQYIAQIOGTQGVDSIDQAGGSQAIIWLNKIPETKG-- 2155
 Qy 1841 VNGEIAPIWNCRLPETTSWNPANPLDAIDPDVAQNDDPMYKTIATFMRLLDQLILRGDM 1900
 Db 2156 -SPLFAPYLSHSTKOLIT-----IAGDI 2179
 Qy 1901 AYRELTRDALNEAKWYVRTLELLGDEPEDYGSQOWAAPSLSGAASQTVQA-AYOQDLTM 1959
 Db 2180 A-----NTLEVI-----ANPNFNDAITNLOINTYTOQMSR 2210
 Qy 1960 LGRGVSKNLTANSVLGLFLPEYNPALTDYQWTLRLFLNLRHLSIDGOPSLSAIYAE 2019
 Db 2211 LAKLSDTSPASA-----DPHE---EALKNKRFAADAIPNAMDVLK 2250
 Qy 2020 PTDFKALITSMVQASQGSQSAVL-PGTLISLYRFPVWLER-TRNLV-----AQL 2064
 Db 2251 YSQRNRKNNVWATGVGGASPFINGGTGLYINVGYDRFIRKGVIVGGVAAVYSGFHHNI 2310
 Qy 2065 TQGTGSLLSMAEHDDA---DELTTLLQOQWELATQSIQORTVDEVDADIADLAESR 2120
 Db 2311 TQSGSNVNMVGYRAFRKRSSELWLSL-----NETWGNKTFINSYDPLLSII-----2358
 Qy 2121 RSAQRNLEKYOQ-LYDEDDINH 2141
 Db 2359 ---NQSKYKDYDTWTDAKINYG 2376

RESULT 15
 T29042
 hypothetical protein B0228.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T29042
 R:Leimbach, D.
 A:Submitted to the EMBL Data Library, March 1995
 A:Description: The sequence of C. elegans cosmid B0228.
 A:Reference number: 218324
 A:Accession: T29042
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-4385 <LEI>
 A:Cross-references: UNIPROT:Q8IG62; UNIPARC:UPI000011018F; EMBL:U23168; PIDN:AAC38807.1.
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Gene: CESP:B0228.3
 A:Introns: 2545/3; 3781/3

Query Match 1.7%; Score 216; DB 2; Length 4385;
 Best Local Similarity 17.8%; Pred No. 0.03;
 Matches 504; Conservative 400; Mismatches 943; Indels 978; Gaps 132;
 Qy 2 YSTAVLLNKISPTRDQGTMTIADLOYSFSELKI-----FDDQLSWGEARHLHYE 52
 Db 1289 WSDISLPEKLQHNISNVVRVLADSSYNCDISIIGKLNAPESQAYVDDQITS-----EHQ 1341
 Qy 53 TIEQKNRLLLEARIFTANQLSCAIRLGIERSDV--SRSYDE-----94
 Db 1342 TLEIVAN-----IKSAVSSVTSNDSVLSKRSDEKAVISNLIISCDL 1386
 Qy 95 -MFGARSSSF-----VKPGSVASNFSPAGVLTLYREAKDLHFSSSYHLDN 140
 Db 1387 ASMSGTSNSFEFORDILEPQANVLLGTIPASQVFTNLEQPIQSEVQ-----GFWSN 1439


```

Db 3103 -----NTGTGKGENASKVLPHKP--PIDTASMKAKAQNSTMTG--SLOKSP--SABA 3150
QY 2058 RNLVAQ-LTOFGTSLLSMAEHDDADELT-TLLLOQGMELATOSIRIQORTVDEVDAIAY 2115
Db 3151 MGIVSQKTEGANSKFGIAQ--GAVETTLTASQSGVTCRDISV-----SNIGV 3197
QY 2116 LAESRRSAQNLEKYOQLYDEIDINHGEOQAMSLDLAAAGQSLAGQVLSIAE-----GVA 2169
Db 3198 ASETVTQPSNK-----ETGIGFCASNLIAPPESAETEFTGKISNLSQTSLNKMAAS 3249
QY 2170 DLVNPVFG-LACGSRWG--AALR--ASASVMSLSATASOVSADKISRSEAYRREROEWE 2224
Db 3250 DVGTTVESKIQAPGNYGDVSLQKVASSDAITKAWQASRDSAISVD----FRQDRE--T 3303
QY 2225 IORDNADGEVKOMDAQLESKI---RREAAQMOV----EYQETOQA--HTQQAQLEL----- 2271
Db 3304 VSAEKSDLNFKSTNSETQKLLFESKEBESGIFIRSSHYEYETQKTLRHSASRESASRT 3363
QY 2272 -----LQKFTNKALYSMMRGKLSAIYYQFFDLTOSFCLMAQEAALRRELT-----DNG 2319
Db 3364 VTAPTNQEVQMNFDKKVEDSVAEGSLSIGIVR--ESSQSEVMQHABR-TSELTKLSMNEE 3420
QY 2320 VTFIR--CGAWNGTTAGLMAGETLLNLAEMEKVWLERDERALEVTRTVSLAQFYQALSS 2377
Db 3421 VAGVRAVSETTNETPRGYQQGDVQVGTGAAMGRIEAPRPG----- 3462
QY 2378 DNFNLTEKLTQFLREGKGNVGASGNELKLSNRQIEASVRLSDLKIFSDYPESLGNTRQK 2437
Db 3463 -EAEITQKLRRTL-----SVERSASAKASMAESQTVTQIQKREDSLASFY--SVRDTLLK 3514
QY 2438 QVSVT 2442
Db 3515 SSSVS 3519

```

Search completed: February 16, 2006, 21:44:50
 Job time : 89.6609 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:43:14 ; Search time 57.1714 Seconds
(without alignments)
3670.210 Million cell updates/sec

Title: US-10-754-115-34

Perfect score: 13043

Sequence: 1 MYSTAVLNKISPTRDGQTM.....KALLESLDIILHIRTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*
1: /cgm2_6/ptodata/1/iaa/5 COMB.pep.*
2: /cgm2_6/ptodata/1/iaa/6 COMB.pep.*
3: /cgm2_6/ptodata/1/iaa/H COMB.pep.*
4: /cgm2_6/ptodata/1/iaa/pCTUS COMB.pep.*
5: /cgm2_6/ptodata/1/iaa/RE COMB.pep.*
6: /cgm2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5148.5	39.5	2522	US-09-251-645-13	Sequence 13, Appl
2	5043	38.7	2516	US-08-851-567B-47	Sequence 47, Appl
3	5043	38.7	2516	US-09-817-514A-2	Sequence 2, Appl
4	4985.5	38.2	2504	US-08-851-567B-12	Sequence 12, Appl
5	4985.5	38.2	2504	US-09-817-514A-8	Sequence 8, Appl
6	3303	25.3	1849	US-08-851-567B-49	Sequence 49, Appl
7	3278.5	25.1	1844	US-08-851-567B-53	Sequence 53, Appl
8	1632	12.5	573	US-08-851-567B-55	Sequence 55, Appl
9	1615	12.4	579	US-08-851-567B-51	Sequence 51, Appl
10	1591	12.2	1565	US-08-851-567B-59	Sequence 59, Appl
11	1553	11.9	1189	US-08-851-567B-36	Sequence 26, Appl
12	1401.5	10.7	845	US-08-851-567B-37	Sequence 37, Appl
13	981	7.5	562	US-08-851-567B-30	Sequence 30, Appl
14	593	4.5	627	US-08-851-567B-28	Sequence 28, Appl
15	429.5	3.3	1095	US-08-851-567B-34	Sequence 34, Appl
16	309	2.4	965	US-08-851-567B-57	Sequence 57, Appl
17	254.5	2.0	10182	US-09-134-001C-3159	Sequence 3159, Ap
18	248.5	1.9	603	US-08-851-567B-35	Sequence 35, Appl
19	246	1.9	5024	US-09-710-279-2964	Sequence 2964, Ap
20	229.5	1.8	2383	US-09-492-709A-302	Sequence 302, App
21	223	1.7	2710	US-08-480-604A-6	Sequence 6, Appl
22	223	1.7	2710	US-08-405-496A-6	Sequence 6, Appl
23	223	1.7	2710	US-08-915-136-6	Sequence 6, Appl
24	223	1.7	2710	US-08-957-310-6	Sequence 6, Appl
25	223	1.7	2710	US-10-011-366-6	Sequence 6, Appl
26	223	1.7	2710	US-09-084-517-6	Sequence 6, Appl
27	200	1.5	2188	US-09-328-352-7763	Sequence 7763, Ap

ALIGNMENTS

RESULT 1

US-09-251-645-13
; Sequence 13, Application US/09251645
; Patent No. 6281413

; GENERAL INFORMATION:

; APPLICANT: Kramer, Vance C.

; APPLICANT: Morgan, Michael K.

; APPLICANT: Anderson, Arne R.

; APPLICANT: Hart, Hope

; APPLICANT: Warren, Gregory W.

; APPLICANT: Dunn, Martha

; APPLICANT: Chen, Jeng S.

; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 2522
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-13

Query Match 39.5%; Score 5148.5; DB 2; Length 2522;

Best Local Similarity 43.0%; Pred. No. 0;

Matches 1121; Conservative 464; Mismatches 822; Indels 197; Gaps 55;

Qy	22	LADLYLSFSLRKIFDDQLSGEARHLYHTIEQKNRLLERARIFTRANPOLSGAIRL	81
Db	28	LTDSHSFNFHQVSEHLSSEADLYHQAQKNRLYEARILKRTNPQLQNAVHL	87
Qy	82	GIERDSVSR-SYDBMFGARSSSVFKPGSVASMPSPAGYLTLYREAKDLHFSSSYHLDN	140
Db	88	AIVAPNAELICYNQFSGRASQYVAPGVTSVMFSPAAVLTLYREARNLHASDSVYRLDT	147
Qy	141	RRPDLADLTLSQNMDETSITLSNELLLEHTIKTGDS-DALMESLSTYRQAITDTPY	199
Db	148	RRPDLKSNALSQNMDETSITLSNELLLEHTIKTGDS-DALMESLSTYRQAITDTPY	207
Qy	200	HQPVTETROVTMTDSTLSALSRNPVQMGASLLAILANISPELYNLTETEETKNA	259
Db	208	HDATENVKVKVQLQDPGLEQLNASPAIAGLMHQSLGINASISPELYNLTETEETKNA	267
Qy	260	DALFAQNFSENITPENFASQSWIARYKYLELSEVQKYLGMQLQ-N-GYSDSTSYAVDNI	318
Db	268	EELYKKNFG-NIEPASLAMPYLRYYNLSDELSQFQIGKASNFQOQVSYN---NQLITP	323
Qy	319	LIVNNEKLEAYKTRVKTDDYDKNINYFD--LMYEGNNQFFIRANFKVSRFGATLRKN	376

28	199.5	1.5	3290	2	US-09-328-352-5486	Sequence 5486, Ap
29	192.5	1.5	1619	2	US-09-328-352-7347	Sequence 7347, Ap
30	192	1.5	2301	2	US-09-822-871-4	Sequence 4, Appl
31	191.5	1.5	2039	2	US-09-077-098A-7	Sequence 7, Appl
32	191.5	1.5	2039	2	US-10-192-584-7	Sequence 7, Appl
33	185.5	1.4	2285	2	US-09-308-375-2	Sequence 2, Appl
34	185.5	1.4	2285	2	US-09-932-183A-2	Sequence 2, Appl
35	185.5	1.4	4536	2	US-09-180-422B-27	Sequence 27, Appl
36	185.5	1.4	4563	2	US-09-108-006C-1	Sequence 1, Appl
37	184	1.4	2172	2	US-09-583-110-5266	Sequence 5266, Ap
38	184	1.4	2238	2	US-09-107-433-4506	Sequence 4506, Ap
39	184	1.4	2777	2	US-09-543-681A-6124	Sequence 6124, Ap
40	182.5	1.4	2366	1	US-08-480-604A-10	Sequence 10, Appl
41	182.5	1.4	2366	1	US-08-405-496A-10	Sequence 10, Appl
42	182.5	1.4	2366	1	US-08-915-136-10	Sequence 10, Appl
43	182.5	1.4	2366	2	US-08-957-310-10	Sequence 10, Appl
44	182.5	1.4	2366	2	US-10-011-366-10	Sequence 10, Appl
45	182.5	1.4	2366	2	US-09-084-517-10	Sequence 10, Appl


```

Qy 2224 EIQRNADGEVKQMDAQLSLKIRREAAQVVEYQETQOAAHTQOALELLOKFTNKALYS 2293
Db 2193 EIQRNAAEALQIDAQKLSAVREAAVLOKTSKTOQEQTSQALPQLOKFSNQALYN 2252
Qy 2284 WMRGKLSAIYYQFDFLTQFCLMAQEAALRRLTONGVTFIRGGAWNGTTAGLMAGETILL 2343
Db 2253 WLRGLAIIYQFYDLAVARCLMAEQAYRWELNDDSAFFIKPGAWQGTIYAGLLAGETIML 2312
Qy 2344 NLAHEKVKWLRDRERALEVTRTVSLAQFYQALSSDN--FNUTEKLTQPLREGKNGVQASG 2401
Db 2313 SLAQMEDAHLKRDRALEVERTVSLAEVYAGLPKONGPFSLAQBIDKLVSQSGSAGSGN 2372
Qy 2402 NELKL-----SNRQIEASVRLSDLKIFSDYPESIGNTRQLKOVSVTLPALVGPVEDIRAV 2456
Db 2373 NNLAFAGAGTDTKTSLOASVSADLKIREDPASLQKIRRIKOISVTLPALLGPPYQDQAI 2432
Qy 2457 LNYGGSIVMPGCGSAIALSHGVNSDQFMLDFNDSRYLPFPGISVNSDGSITLSFPDPAT- 2515
Db 2433 LSYDGKAGLANGCEALAVSHGWNDSGQFQDFNDGKFLPFEGIAL-DOGTILTSFPNASM 2491
Qy 2516 ---DRQKALLESLSDIILHIRYTR 2537
Db 2492 PEKGQATMLKTLNDIILHIRYTIK 2516

```

RESULT 4

```

US-08-851-567B-12
; Sequence 12, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Pettig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merio, Donald J.
; APPLICANT: Ori, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851.567B
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063.615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395.497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007.255

```

```

; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608.423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705.484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-851-567B-12

Query Match 38.2%; Score 4986.5; DB 2; Length 2504;
Best Local Similarity 42.3%; Pred. No. 0;
Matches 1110; Conservative 434; Mismatches 863; Indels 217; Gaps 48;

Qy 9 NKISPTDGG-----QWTMLADQLVSPSELKIFDDQLSWGEARHLTHETIEOKKNRL 63
Db 3 NSLSSTDTTICQKQLTCPAEIALYPEDTFRKTRGMVNWGEAKRIYEIAQAEQDRNLH 62
Qy 64 EARIFRANPOLSGAIRLGIERD-SVSRSYDEMFGARSSSVKVPGSVASFSPAGYLTEL 122
Db 63 EKRIFAYANPLLNKAVRLGTRQMLGFIQGSVDLFGNADYAAPGSVASFSPAYITEL 122
Qy 123 YREAKDLHFSSAYHLDNRRLADLTLQSNDMTEISTLTLSNELLEHITRTKGDSD 182
Db 123 YREAKNLHDSSIVYLDKRRPDLASMLSQKNDEEISTLALSNELCLAGIETTKGSQD 182
Qy 183 ALMESLSTYQADITPYHQPYETIRQVIMTHDSTLSALSRNPEVMGQAGASILLAIANI 242
Db 183 EVMDMLSTYRLSGETPYHHAYETVREIVHERDPGRHLSQAPIAAKLDPVTLGISSHI 242
Qy 243 SPELYNLTBEITEKN---ADALFAQNFSENIIPENFASQSWIAKYGLSEVQKYLGM 299
Db 243 SPELYNLTBEITEKN---ADALFAQNFSENIIPENFASQSWIAKYGLSEVQKYLGM 299
Qy 300 LQN-GYSDSTSAYVDNI-STGLVNNESKLBAYKITRVKTDYDKNINIFDLMEYGNQOFF 358
Db 302 LSHVGYG-----SDILVPLVDGVGKMEVVRVTRTPSDNYTSQTYIELYPQGDNYL 354
Qy 359 IRANFKVSRFEGAT---LRKNAGPSGIVGSLGP---LIANTNPKSNVLSNISSEYKNG 412
Db 355 IKYN--LSNSFGLDDFYLYQKQSGADWTEAHNPYPDMVINQYKYESOATIKRSDSD--NI 410
Qy 413 VKIYAVRYTSTGATNQGIGFTFESYPLTIFALKLNKAIKRLCLTSLGSLNELOTVRSD 472
Db 411 LSLGLQWHS--GSYNFAANFKIDQSPKAFLLKMKKALRLKATGLSATLERIVDSV 468
Qy 473 NAOGIINDSVLTKVFTLYSHRYALSFDDAQVINGSVINQYADDDSVSHNFRNFTNPL 532
Db 469 NSTKSIITVELNKVYRVKVIDRYGISEETAAILANINISQAVGNQLSQPEQLFNHPPL 528
Qy 533 KGKIFE-ADGNT-----VSDPDEQSTFASALMRGLGVNSGELYOL-----GKLAG 579
Db 529 NGIRYEISEDNSKHLNPNPDLNLKPDSTGDDOKAVLKRAFOVNASELYQMLLITDRKEDG 588
Qy 580 VL--DAQNTITLSVFVISSLYRLTLARVHOLTWNELCMYGLSPFNGKTTASLSGELP 637
Db 589 VIKNNLEN-----LSDLYLSLLAQINLITAEINILLVICGYGDTNLYQITDNL 640
Qy 638 RLVIWLYQVTOWLTEABITTEALIWLLCTPFGSNGISPEISNL-----LNNRPSISED 690
Db 641 KIVETLLWITOWLKTQKVTVDLFLMTATYTTTLTPEISNLTATLSSTLHGKESLIGED 700

```



```

QY 1960 LGRGVSKN--LRTANSLVGLFPEYNPALTDYWTQLRLRLNLRNLSIDQPLSLAIY 2017
Db 1921 LRLSRVKTPLLTGANSALTALFLPQENSKLKGWRTLAQRMFNLRNLSIDQPLSLPLY 1980
QY 2018 ABPTDPKALLTSWQASOGGSNAVLPGTSLRFPVWLETRNLVAQLTQFGTSLLSMAEH 2077
Db 1981 AKPADPKALLSAVSGQGDLPKAPLTIHRRPQMLEGARGVLNQLIQFGSSLLGYSER 2040
QY 2078 DDADBLTLLLOQGMELATQSIROQRTVDEVDADIAVLAEARRSAQNRLEKYQOQYD 2137
Db 2041 QDAEAMSQLQTOASELILTSIRMQDNQLAELDSKTAQVSLAGVQQRFPDSYSLYEN 2100
QY 2138 INHGQRAMSLDAAAGOSLAGQVLSIAEGVADLVNPNVFLGACGSRWGAALRASAVMS 2197
Db 2101 INAGEORALALRSESAIESQGAQISRMAGAGVDMAPNIFGLADGGMHYGATAYALADGIE 2160
QY 2198 LSATASQYSADKISSEAYRRRQWEIORDNADCEVKMDAQLESKIRREAAQOVY 2257
Db 2161 LSASAKMVDAAEKVAQSEYRRRQWEIORDNADCEVKMDAQLESKIRREAAQOVY 2220
QY 2258 OETQOAHQAOLELQRTFKALYSWMRGKLSAIYYOFFDQTSQFCILMAQOALRRELTD 2317
Db 2221 LKTOQAQAQUTLRSFESQALYSMLRGLSGIYFQYDLAVSRCLMAEQSYQWEAND 2280
QY 2318 NGVTFIRGAMNGTTAGLMAQETLLNLAEMKWLDERDERALEVTRTVSLAQFYQAL-S 2376
Db 2281 NSISVPKGMQGTVAGLLCGEALIQNALQEAELKWSRALEVERTVSLAVVYDSLEG 2340
QY 2377 SDNFNLTEKLTQFLREGKNGVAGSNELKSNRQIEASVRLSDIKIPSDYPSL---GNT 2433
Db 2341 NDRFNLAEQIPALLDKGEGTAGTKENGLSLANAILSASVKLSDLKLTDPDSVGSNKV 2400
QY 2434 ROLKQVSVTLPALVGPYDIRAVLNYGGSIVMPCGSAIALSHGVNDGQFQMLDFNDSRY 2493
Db 2401 RRIKQISVSLVGPYQDVQMLSGYSGSTQLPKGCASALAVSHGNDGQFQMLDFNDSRY 2460
QY 2494 LPFEGISVNDGSLTSPFPDTRDQKALLESIDILHRYTIR 2537
Db 2461 LPFEGIALDDQDTNLQFNPNATDKQKAILQTMDSIILHRYTIR 2504

```

RESULT 6

```

US-08-851-567B-49
; Sequence 49, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: French-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapingda, Kitisiri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dow AgroSciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,567B
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1849 amino acids
; TYPE: amino acids
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-851-567B-49

```

```

Query Match 25.3%; Score 3303; DB 2; Length 1849;
Best Local Similarity 39.3%; Pred. No. 1.2e-252;
Matches 764; Conservative 339; Mismatches 658; Indels 182; Gaps 51;

QY 92 YDEMFGARSSFFVKGVSASHFSPAGYLTLYREAKDLHFSSAYHLNRRPDLADLTLS 151
Db 4 YNNQFSGRASQYVAPGTVSSMFSPPAAYLTLYREARNLHASDSVYLLDTRRPDLKSMALS 63
QY 152 QSNMDTEITLTLNELLLEHI-TRKTGSDSALMESLTYRQAITDPVHQPYETIROVI 210
Db 64 QONMDIELTUSLSNELLESIKTESKLENTTKVMEMLSTFPSPGATPYHDAYENREVI 123
QY 211 MTHDSTLSALSRNPEVMQAGSALLAILANISPELYNLTETETKKNADALFAQNFSEN 270
Db 124 QLQDPGLEQLNASPAIAGLHQAQSLGINASISPELFNLTETETEGNAEELYKKNFG-N 182
QY 271 ITPENFASQSWIAKYIGLESEVKYLGMLQN-GYSDSTSAVDNISGLVNNESKLEA 329
Db 183 IEPASLAMPYELKRYNLSDEELSQFIGKASNFQGOEYSN---NQLITPVVNSSDGTVKV 239
QY 330 YKTR-VKTDYDKNINVDLMYEGNNOFFFRANFKVSREFGAT-----LRKN 376
Db 240 YRITREYTTNAYQMDVELFP--FGGEN--YALDYKFNFNYNASVLSIKLNDKRELVRTE 294
QY 377 AGPSGIVGSLSGPLIANTFNKSNVLSNISDSSEYKNGVKIYAYRYTSTSATNOGGGIFTF 436
Db 295 GAQP-----VNIEYSANITLTAD--ISQPFBIGLTRVLPSPGSA-YAAAKFTV 340
QY 437 ESYPLTIPALKKLNKAIKRLCLTSGLSPNELQTVIRSDNAQGIINDSVLTKVFTLYSHRY 496
Db 341 EBYNQYSPLLKLNKAIKRLSRATELSPTILEGIVRSVNLQDINTDVLGKVFLLTKYMQRY 400
QY 497 ALSFDQAQVANGSVINQYADDSDSVSHFNRLFNTPKPKGIPEADGNTVTSIDPDEQSTFA 556
Db 401 AIHAETALILCNAPISQBSYDNQPSQFQRLFTPLNGQYFSTGDEEIDLN-SGSTGDWR 459

```


Db 1636 KDPHGFSLDDHKTFSGLSQAQKN-----DSEPMDFSGANALYFWELP 1680
Qy 1807 YITPMCFQRLLQSKQDEATQWYVYVNGEIPWICRPLEBETTSWNP 1866
Db 1681 YITPMMAHRLQBNFQFADANHWFRYVWSPSGYVDGKIATYHWNVRPLEEDTSSNAQOL 1740
Qy 1867 DAIDPDVAQNDPMHYKIATPMRLLDQILRGDMAYRELTFDALNEAKMNVVRLLELGD 1926
Db 1741 DSTDPDAQDDPMHYKVATPMATLDLLMARGDAAYRQLERDTLAEAKMNTQALNLGD 1800
Qy 1927 EPEDYGSQQAAPSLGSAASQVQAAYQQDLTML 1960
Db 1801 EPQVWLSTTWANPFLGNAASKTTQVRAQVLTQL 1834

RESULT 8

US-08-851-567B-55
; Sequence 55, Application US/08851567B
; Patent No. 6528484

GENERAL INFORMATION:

; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: fFrench-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitirai

TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus

NUMBER OF SEQUENCES: 88

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Dow Agrosciences Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US

ZIP: 46268

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/851.567B

FILING DATE: 05-MAY-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/063,615

FILING DATE: 18-MAY-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/395,497

FILING DATE: 28-FEB-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/007,255

FILING DATE: 06-NOV-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/608,423

FILING DATE: 28-FEB-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/705,484

FILING DATE: 28-AUG-1996

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J

REGISTRATION NUMBER: 27386

; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 608-251-5000

; TELEFAX: 608-251-9166

; INFORMATION FOR SEQ ID NO: 55:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 573 amino acids

; TYPE: amino acids

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-851-567B-55

Query Match

12.5%; Score 1632; DB 2; Length 573;

Best Local Similarity 55.3%; Pred. No. 1.3e-120;

Matches 317; Conservative 105; Mismatches 147; Indels 4; Gaps 2;

Qy 1969 LRTANSLVGLFPEYNPALTDYMTLRRLNLFNLRHNLSDIQPLSLAIYAEPTDPKALLT 2028
Db 1 LGTANSLTALFLPQENSKLKGWRTTLAQMFLNRHNLSDIQPLSLPLYAKPADPKALLS 60
Qy 2029 SMYQASQCGSAVLPGTSLSYRPPVWLERTNLVAQLTQFGTSLLSMAEHDDADLTLL 2088
Db 61 AAVSASQGGADLPKAPLTIHFPQMLEGARGLVNQLIQFGSSLLGYSERQDAEAMSQLQ 120
Qy 2089 QCGMELATQIRIQORTVDEVDADIAVLBSRRSAQNRLEKYQOLYDDEINHGQRAMSL 2148
Db 121 TQASELILTSIRMQDNLAEDESEKTAQLVSLAGVQORFDSYSQLYEENINAGQRALAL 180
Qy 2149 LDMAGQSLAGQVLSIAEGVADLVPNVFGACGSRMGAAALRASAVMSLSATASQYSAD 2208
Db 181 RSESAIESQGAQISRAGAGVDMAPNIFGLADGGMHYGATAYATADGIELSASAKWDAE 240
Qy 2209 KISRSEAYRRRROWEIQRDNAGEVQMDAQLESKIRREAAQOVVEYQETQQAHTQAAQ 2268
Db 241 KVAQSHYRRRRQEWKIQRDNAQAEINQLNAQLESLSIRREAAEMQKEYLKTQQAQAAQ 300
Qy 2269 LEQLQRFTKALYSVMRGKLSAIYQFDFLTQSFCLMAQOELRRELTNGVTFIRGAW 2328
Db 301 LTFLRSKFSNQLYSWLRGLSGIYFQFDLAVSRCLMAEQSYQWEANDNSISFVKFGAW 360
Qy 2329 NGTTAGLMAGETLILNLAEKMWLERDERALEVTRTVSLAQFYQAL-SSDNFNLTKLT 2387
Db 361 QGTVAGLLCGEALIQNLAEAYLKWESRALEVERTVSLAVVYDSLEGNDRFNLAEQIP 420
Qy 2388 QFLREGKNGVAGSNELKLSNRQIEASVRLSDLKIFSDYPESL---GNTRQLKQVSVTL 2444
Db 421 ALLDKGEGTAGTKENGSLANAILASVYKLSDLKLTGTDYDPSIVGSKNVRRIKQISVSL 480
Qy 2445 ALVGPYEDIRAVLNYSIVMPRGCSAIALSHGVNDSGQFMLDFNDSKYLFPFGISVND 2504
Db 481 ALVGPYQDVQAMLSYGGSTQLPKGCSALAVSHGVNDSGQFQLDNFGKYLFPFGIALDDQ 540
Qy 2505 GSLTSPFPDATDROKALLESLSLIIHRYTIR 2537
Db 541 GTLNLQFPNATDKQKAILQTMSDIILHRYTIR 573

RESULT 9

US-08-851-567B-51

; Sequence 51, Application US/08851567B

; Patent No. 6528484

; GENERAL INFORMATION:

; APPLICANT: Ensign, Jerald C

; APPLICANT: Bowen, David J

; APPLICANT: Petell, James

; APPLICANT: Fatig, Raymond

; APPLICANT: Schoonover, Sue

; APPLICANT: fFrench-Constant, Richard

; APPLICANT: Rocheleau, Thomas A.

; APPLICANT: Blackburn, Michael B.

; APPLICANT: Hey, Timothy D.

; APPLICANT: Merlo, Donald J.

APPLICANT: Orr, Gregory L.
APPLICANT: Roberts, Jean L.
APPLICANT: Strickland, James A.
APPLICANT: Guo, Lining
APPLICANT: Ciche, Todd A.
APPLICANT: Sukhapinda, Kitisri
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESS: Dow AgroSciences Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,567B
FILING DATE: 05-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255
FILING DATE: 06-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,423
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/705,484
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.93804
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 579 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-851-567B-51

Query Match 12.4%; Score 1615; DB 2; Length 579;
Best Local Similarity 54.5%; Pred. No. 3.1e-119;
Matches 316; Conservative 108; Mismatches 144; Indels 12; Gaps 4;

QY 1969 LRTANSLVGLPEVNPALTDYQTLRLRLFNRLNLSIDGQPLSLAIYASPTPKALLT 2028
DB 1 LRSANTDLFLPQINVMYMYWTLAQRVNRLNLSIDGQPLSLAIYATPADPKALLS 60

QY 2029 SMVQASQGSVALPGTSLXFPFVLMRLNVAQLTQFTSLISMASHDDADELTLLLL 2088
DB 61 AAVATSGGGKLPESFMSLWPPHMLNARGVSLQTFGSTLQNIIEQDAEALNALLQ 120

QY 2089 QQGMELATQSIQRTQVDEVDADIAVLAEGRSSAQNRLKLYQQLYDEIDNHGEORAMSL 2148
DB 121 NQAAELILNLSIQDKTEELDAEKTLEKSKAGQSRFDSYGLYDENINAGENOAMTL 180

QY 2149 LDAAGQSLAGQVLISAGVADLVNPFVGLACGSRWCAALRASASVMSLSATASQYSAD 2208
DB 181 RASAGLTAVQASRLAGAAADLVNIFGPFAGGSRWGAIAEATGYVMEFSANVMNTAD 240

QY 2209 KIRSEAYRRRRQEWIQRDNADGEVKQMDAQLESKLIRBAAQMQVEYQETQQAHQAO 2268
DB KISQSEYRRRRQEWIQRDNADGEVKQMDAQLESKLIRBAAQMQVEYQETQQAHQAO 300

QY 2269 LELQKRTKALYSWMRGKLSAIYYQFDDLTQSFCLMAQALRELTDNGVTFTRGGAW 2328
DB LAFLQKFSNALYNLWLRGLAAIYFYFDLAVARCLMAEQAYRWELNDDSAARFKPGAW 360

QY 2329 NGTTAGLMAGETLLNLNLAEMEKVWLERDERALEVTRTVSLAQFYQALSSDN--FNLTKL 2386
DB QGTYAGLLAGETLLMLSLAQMDAHLKRDRALEVERTVSLAEVYAGLVKPDNGPFLADEI 420

QY 2387 TQFLREGKNGVSGSNELKL-----SNRQIEASVRLSLDKIFSDYPESLGNTRQLKQVSV 2441
DB DKLVSQGSAGSGNNLAFAGAGTDTKTSLQASVSFADLKIREDPYASLGKIRRIKQISV 480

QY 2442 TLPALVGYEDIRAVLNYGGSIVMPRGCSATALSHGVNDSQFMDLDFNDSRYLPEGISV 2501
DB TLPALLGYQDVQATLSYDGRAGLANGCEALAVSHGMNDSQGFQDFDNDGKFLPPEGIAI 540

QY 2502 NDSGLTSLSPDAT----DROKALLESLSDIILHRYTIR 2537
DB -DQGTILSLFNFNASMPKKGQATMLKTLNDIILHRYTIK 579

RESULT 10
US-08-851-567B-59
Sequence 59, Application US/08851567B
Patent No. 6528484
GENERAL INFORMATION:
APPLICANT: Ensign, Jerald C
APPLICANT: Bowen, David J
APPLICANT: Petell, James
APPLICANT: Fatis, Raymond
APPLICANT: Schoonover, Sue
APPLICANT: French-Constant, Richard
APPLICANT: Rocheleau, Thomas A.
APPLICANT: Blackburn, Michael B.
APPLICANT: Hey, Timothy D.
APPLICANT: Merlo, Donald J.
APPLICANT: Orr, Gregory L.
APPLICANT: Roberts, Jean L.
APPLICANT: Strickland, James A.
APPLICANT: Guo, Lining
APPLICANT: Ciche, Todd A.
APPLICANT: Sukhapinda, Kitisri
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESS: Dow AgroSciences Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,567B
FILING DATE: 05-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255

RESULT 11
 US-08-851-567B-26
 ; Sequence 26, Application US/08851567B
 ; Patent No. 6528484
 ; GENERAL INFORMATION:
 ; APPLICANT: Ensign, Jerald C
 ; APPLICANT: Bowen, David J
 ; APPLICANT: Pettit, James
 ; APPLICANT: Patig, Raymond
 ; APPLICANT: Schoonover, Sue
 ; APPLICANT: French-Constant, Richard
 ; APPLICANT: Rocheleau, Thomas A.
 ; APPLICANT: Blackburn, Michael B.
 ; APPLICANT: Hey, Timothy D.
 ; APPLICANT: Merlo, Donald J.
 ; APPLICANT: Orr, Gregory L.
 ; APPLICANT: Roberts, Jean L.
 ; APPLICANT: Strickland, James A.
 ; APPLICANT: Guo, Lining
 ; APPLICANT: Ciche, Todd A.
 ; APPLICANT: Sukhapinda, Kitisri
 ; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
 ; NUMBER OF SEQUENCES: 88
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Dow Agrosciences Patent Department
 ; STREET: 9330 Zionsville Road
 ; CITY: Indianapolis
 ; STATE: IN
 ; COUNTRY: US
 ; ZIP: 46268
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/851,567B
 ; FILING DATE: 05-MAY-1997
 ; CLASSIFICATION: 514
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/063,615
 ; FILING DATE: 18-MAY-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/395,497
 ; FILING DATE: 28-FEB-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 60/007,255
 ; FILING DATE: 06-NOV-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/608,423
 ; FILING DATE: 28-FEB-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/705,484
 ; FILING DATE: 28-AUG-1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Seay, Nicholas J
 ; REGISTRATION NUMBER: 27386
 ; REFERENCE/DOCKET NUMBER: 960296.93804
 ; TELEPHONE: 608-251-5000
 ; TELEFAX: 608-251-9166
 ; INFORMATION FOR SEQ ID NO: 26:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1189 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-851-567B-26

Query Match 11.9%; Score 1553; DB 2; Length 1189;
 Best Local Similarity 27.0%; Pred. No. 9.9e-114;
 Matches 447; Conservative 225; Mismatches 477; Indels 504; Gaps 45;

QY	916	QTLAENMEAGLSTQQAOTFLADYTAERLSSVLGNWFLANIQPEGVSLHRRDDLYSYFLIDN	975
DB	8	QTLKEARDAL-----VAHYIATQVPADLKE-----SIQTADDLYEYLLD	49
QY	976	QVSSAIKTRLAELAGIQLYNRAL-----NRIENARADYSTROFFFTDM-TVNNRYSTW	1030
DB	50	KISDLVTTSPLEAIGSLQFLHRAIEGYDGTLDASAKPYFADEQFLWNDSFNHRYSTW	109
QY	1031	GGVSRVYYPENYIDPTORIGQTRMMDELLENISQSLSRDTEVEFAKTYLTFRTFVADL	1090
DB	110	AGKERLKEYAGDYIDPTLRANKTEIFTAFEGISQGLKSELVESKLRDLYLSYDTLATL	169
QY	1091	KVVSAYHDNVSNTGLTWFGQTRLENLPEYVNRVNDISRMQAAGELAAANAKWETKIDTAV	1150
DB	170	DYITACQCKDKNT---IFFIGRTQNAFYAWRKLTLV-TDGGKLPQWSEWRAINAGI	225
QY	1151	N-PYKDAIRPVI FRERLHLI WKEEVAKNGTDPVETVYDRFTLKLAFLRHDSWSAPWSY	1209
DB	226	SEAYSGHVEFPFENNKLHIRM-----FTI-----	249
QY	1210	DITTOVEAVTDKPDTERLALAAAGFQGEDTLLVFFVYKTKSYSDFGGSKNKNVAGMTI-Y	1268
DB	250	-----SKEDKIDFVY-----KNIWMSSDY	269
QY	1269	GGGFKKMENTALSRYSQLKNTFDIIHTQGNLVRKASYRFAQDFEVPAASLNGSAIGDD	1328
DB	270	SWASKKILELSFTDYNRVGAT-----GSSSPTEVASQYGS-----AQMNISD-----DG	315
QY	1329	SLTWEN--GNIPOITSKYSSDNLAITLHNAAFTRYDGSNVIIRNKQISAMKLTGVGDK	1386
DB	316	TVLIFQAGGATP-----STGVTLCYD-SGNVIKN-----L	345
QY	1387	SOYGNAFIANTVKHYGYSGLGGPITVYNKTKNYIASVQGHLMNADYTRRLILTEVENN	1446
DB	346	SSTGSANLSSKD-----YATTK-LRMCHGQSYN-----DNN	375
QY	1447	YYARLFEPFPSPNTILNTFTVGSNKTSDFKKCSYAVDGNNSQGQPIFSSVQSSGWLID	1506
DB	376	Y-----CNFTLSINTI-----EFTSY-----	391
QY	1507	TGINNTDIKITVMAGSKTHTTASDHIAASLPANSFDAMPYTFKPLEIDASSLAFTNNIAP	1566
DB	392	-----GTSSDGKQFTPPSG-----	407
QY	1567	LDIVFETKADGRVLGKIKQTLVSKRVNYPEDILFLRETHSGAQYLMQGVYRIRLNTILL	1626
DB	408	-----AIDLHLFNY-----VDLNALL	423
QY	1627	ASQLVSRANTGIDTILTWETQRLPEPPLGEGFFANFLPKYDPAEHGDERWFKIHGNVG	1686
DB	424	-----DISLDSLLNYDVO-----GQPG	440
QY	1687	GNTGQPYSGMLSDTSETMTLFPVYAEGYMHGVLGVGVQKITYDNTWESAFFYD	1746
DB	441	G-----	441
QY	1747	ETKQOFLINDADHDSGMTQOQIVKNIKKYKGLNVSIAATGYSAPMD-FNSASALYYWEL	1805
DB	442	-----SNPVDNFGPYGIYLWEI	459
QY	1806	FYYTPMCFQRLLOKQFDEATQWYNYNYPAGY-IVNGEI-----APWINCRPLEETT	1859
DB	460	FFHIPFLVTRMQTEORYEDADTWYKIFRSAGYRDANGQLIMDGSKPKRYWNVPLQD	519
QY	1860	SWANPLDAIDPDVAQNDPMHYKATFWRLDQLILRGDMAYRELTDALNEAKWYVR	1919
DB	520	AWDTTQPATTPDVIAMADPMHYKLAIFLHTLDLLIARGDSAYRQLERDITLVEAKMYIQ	579
QY	1920	TLELLGDEPEDYGSQOWAAPSLS---GAASQTVQAAQOQDLTM---LGRGGVSKNLRTAN	1973
DB	580	AQQLLGPDPDHTTNTWNPFLSKEGAIAITFTLSSPEVMTFAAWLSAGD-----TAN	633

1974	Qy	SLVGLFPEYNPAITDYWQTLRLRLFNLRNLSIDGQPLSLAIYAEPDPKALLTSMVOA	2033
634	Db	IGGDFFPPYNDVLLGYWDLKLELRNLRNLSIDGQPLNPLIYATVPDPKTL--ORQOA	691
2034	Qy	SOQSSAVLP---GTLSLYFPVWLETRNLVAQLTQFGTSLLSMAEHDDADELTWLLQ	2089
692	Db	GGDGTGSSPAGQSGVQGWRYPLILVERKARSVLLTQFGNSLQTTLEHQDNEKWTILLQT	751
2090	Qy	QCMELATQTSIRIQORTVDEVDADIAVLAESSRSAQNRLKEYQQLYDEDINHGEQRAMSL	2149
752	Db	QOEAILKHQHDIQQNNLKGLOHSITALQASRDGDTLRQKHYSDLINGGLSAAEIAGLTLR	811
2150	Qy	DAAGOSLAGOVLSIABGVADLVPNVFLGACGSRGWAALRASASVMSLSATASQVSAOK	2209
812	Db	STAMITNGVATGLLIAGGIANAVPNVFLANGCSWEGAPLIGSGQATOVGCAGIQDOSAGI	871
2210	Qy	ISRSEAYRRROEWEIORDNADGEVQMDAQLBSLXIKIRREAAQOVVEYQTOQAHTOACL	2269
872	Db	SEVTAGYQRROEBAWALQORDIADNEITQLDAQIOSLQEQITMAQKQITLSETEQANAQAIY	931
2270	Qy	ELLQRKFTNALYSWMRGKLSAIYYOFFDLTQSFCLMAQBALREBLTD--NGVFFIRGGAW	2328
932	Db	DLQTRTFQALLYNWAGRLSALIYYQWYDSTLPICLQPKAALVQELGEXESDSLFPVW	991
2329	Qy	NGTTAGLMAGETLLLNLAEMEKWLERDERALEVTRVTSIAQYQALSSDNFNLTKELTQ	2388
992	Db	NDLWQGLLAGEGLSSELOKLDAILWARGGTGLEBAIRTVSLDTLTF----GTGFLSENINK	1046
2389	Qy	FLREGKGNVCASGNELKLSNRQIEASVRLSDLKIPSDYPESLGN--TROLKQVSVTLPAL	2446
1047	Db	VL--NGETVSPSGGVTLALTGDIFQATLDLSQLGDMSY--NLGNKKRRIRKIRAVTLPTL	1103
2447	Qy	VGPVEDIRAVLNYGGSIIVMPRGCSAIALSHGVNDVDSQFMDFDNDRYLFPFGISVNDSDS	2506
1104	Db	LGPYQDLAEALVMAEATA-----ALSHGVNDGGREFVTFDNDSFLEPPEGEDAT--TGT	1154
2507	Qy	LTSFPPDA--TDRQKALLESLSIILHIVRTIR	2537
1155	Db	LEANIHPAGKEGTQHELVAHLSIIIVHLNVIIR	1187

RESULT 12

US-08-851-567B-37

? Sequence 37, Application US/08851567B

? Patent No. 6528484

? GENERAL INFORMATION:

? APPLICANT: Ensign, Gerald C

? APPLICANT: Bowen, David J

? APPLICANT: Petell, James

? APPLICANT: Fatig, Raymond

? APPLICANT: Schoonover, Sue

? APPLICANT: Ifrench-Constant, Richard

? APPLICANT: Rocheleau, Thomas A.

? APPLICANT: Blackburn, Michael B.

? APPLICANT: Hey, Timothy D.

? APPLICANT: Merlo, Donald J.

? APPLICANT: Orr, Gregory L.

? APPLICANT: Roberts, Jean L.

? APPLICANT: Strickland, James A.

? APPLICANT: Guo, Lining

? APPLICANT: Ciche, Todd A.

? APPLICANT: Sukhapinda, Kitiari

? TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus

? NUMBER OF SEQUENCES: 88

? CORRESPONDENCE ADDRESS:

? ADDRESSEE: Dow Agrosciences Patent Department

? STREET: 9330 Zionsville Road

? CITY: Indianapolis

? STATE: IN

? COUNTRY: US

? ZIP: 46268

? COMPUTER READABLE FORM:

? MEDIUM TYPE: Floppy disk

QY 1017 PTDH-TVNNRSTWGGVSRVLYPENYIDPTORIGQOTMMDELLLENISQSKLSRDTVED 1075
Db 465 FIDWDKYNKRYSTWAGVSQVLYPENYIDPTWRIGQOTKQMDALLQSVSQSLNADTVED 524
QY 1076 APTKYLTRTETVADLKVVSAHDNVNSNTGLTFVQOTRENLPYVYWRNVDIRMQAGEL 1135
Db 525 APMSTLTSPEQVANUKVISAHDNNNDQGLTFYGLSETDAGEYWRVSRVDSKFNQDGP 584
QY 1136 AANAKWEMTKIDTAVNPYKDAIRPVIIFRERLHLIWEKEEVAK---NGTD--PVETDYDF 1190
Db 585 AANANSEWHKIDCPINPYKSTIRPVIYKSLRYLLMLEQKEITKQTSNKGQVQOTEDYR 644
QY 1191 TLKLAFLRHGWSAPWSVIDITTOVEAVTDKPDPTERIALAASQFGQEDTLLVFKYTKG 1250
Db 645 ELKLAHRYDGTWNTPTITPDVNNKISEL--KLEKNRAPGLYCAGYQGEDTLLVFMFYNOQ 702
QY 1251 SYSDFGGSKNVAGTIYGDGSPKQMENTALSRYSLQNTPTDI IHTQGNDLVRKASYRPA 1310
Db 703 TLDSY--KNASQGLYIFADMAKMDTPQSNVYRD--NSYQOFTNN---VRRVNNRYA 755
QY 1311 ODPEVPASLNMGS--AIGDSSLTVMENGNIPOITSKYSNDLAI-----TLHNAAFV 1361
Db 756 EDEYEPSSVSRKDYGMGVDYLSVMYNGDIPITNYKAASSDLKIYISPKLRIHNG---- 811
QY 1362 RYDGSNVIRNKQISAMKLTGVDGSKSQYNAFIANTVVKHYGGYSDLG 1409
Db 812 -YEGQ---KRNQCNLMNKYG-----KLGDKEIV-----YTSLG 840

RESULT 13

US-08-851-567B-30
; Sequence 30, Application US/08851567B
; Patent No. 6528484
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Patig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: French-Constant, Richard
; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Ciche, Todd A.
; APPLICANT: Sukhapinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dow Agrosciences Patent Department
; STREET: 9330 Zionaville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/851,567B
; FILING DATE: 05-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497

; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 562 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-851-567B-30
Query Match 7.5%; Score 981; DB 2; Length 562;
Best Local Similarity 40.1%; Pred. No. 8.3e-69;
Matches 231; Conservative 94; Mismatches 223; Indels 28; Gaps 10;
QY 1971 TANSVLGLFLPEYNPALTDYQWTLRLRLFNLRHNSIDQPLSLAIYABPTDPKALLTSM 2030
Db 4 TANIGDGDPLPPYNDVLGLYWDKLELRNLNLRHNLSDGQPLNPLPYATPVPDKTL--OR 61
QY 2031 VOASQGSAAVLP---GTLSLYRPPVLMLETRNLVAQLTOFGTSLLSMAEHDDADELTTL 2086
Db 62 QOAGDGTGSSPAGGQSGVQGWRYPLVERARSVSLTQFNSLQTTLEHQDNKMTLL 121
QY 2087 LLOQGMELATQSIRIQRTVDEVDADI AVLAEARRSAQNRLKRYQQLYDEIDINHGEQRAM 2146
Db 122 LQTQEAELKHQHDIQNNLKLQSLTALQASRDGDLTRQKHYSDDLINGLSAAEAGL 181
QY 2147 SILDDAAQSLAGQVLSIABGVADLVNVPVFLGACGSGRWGAALRASVMSLSATASQYS 2206
Db 182 TLRSMTITNGVATGLLIAGGIANAVPNVPFLANGSGSEWGAFLIGSGQATQVAGIQDQS 241
QY 2207 ADKISRSEAYRRRQEWEIFORDNADGEVKMDAQLSKIRREAAQMOVVEQTOQAHTQ 2266
Db 242 AGISEVTAGYQRQEEWALQRIADNEITQDQAQISLQEQEITMAQKQITLSETQANQ 301
QY 2267 AQLELLQRKFTNKALYSMMRGKLSAIYYQFFDLTQSFCLMAQEAALRELTLD-NGVTFIRG 2325
Db 302 AIYDLQTTFTGTQALYNNMAGRLSALYYQMYDSTLPICLQPKAALVQELGEKESDLSFOV 361
QY 2326 GAWNGTTAGLMAGETLLNLNLAEMEKVLEDERALEVTRTVSLAQFYQALSSDNFNLTOK 2385
Db 362 PWNLDLWQGLLAGEGLSSELQKLDALNWLARGGICGLEAIRTVSLDTLF-----GTGTLSEN 416
QY 2386 LTQFLREGKNGVGSAGNELKLSNRQIEASVRLSDLKIFSDYPESIGN--TRQKQVSVTL 2443
Db 417 INKVL-NGETVSPSGGVTLALTGDIFQATLDSQLGLONSF--NLGNKKRIKRIAVTL 473
QY 2444 PALVGPYEDIRAVLNLYGSGIVMPRGCSAIALSHGVNDSQGMFLDFNDSRYLPFFEGISVND 2503
Db 474 PTLGLPYQDLEATLVMGAEIA-----ALSHGVNDGGRFVTFDFNDSRFLPFEGRDAT- 524
QY 2504 SGSITLSPDDA--TDRQKALLESIDIIHLHRYTIR 2537
Db 525 TGTLELNIFHAGKEGTQHELVAIVLSDIIVHLNIIR 560
RESULT 14
US-08-851-567B-28
; Sequence 28, Application US/08851567B

Patent No. 6528484
GENERAL INFORMATION:
APPLICANT: Ensign, Jerald C
APPLICANT: Bowen, David J
APPLICANT: Petell, James
APPLICANT: Fatig, Raymond
APPLICANT: Schoonover, Sue
APPLICANT: French-Constant, Richard
APPLICANT: Rocheleau, Thomas A.
APPLICANT: Blackburn, Michael B.
APPLICANT: Hey, Timothy D.
APPLICANT: Merlo, Donald J.
APPLICANT: Orr, Gregory L.
APPLICANT: Roberts, Jean L.
APPLICANT: Strickland, James A.
APPLICANT: Guo, Lining
APPLICANT: Ciche, Todd A.
APPLICANT: Sukhapinda, Kitisiri
TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESS: Dow Agrosciences Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/851,567B
FILING DATE: 05-MAY-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255
FILING DATE: 06-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,423
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/705,484
FILING DATE: 28-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.93804
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 627 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-851-567B-28
Query Match 4.5%; Score 593; DB 2; Length 627;
Best Local Similarity 20.3%; Fred. No. 7.7e-38;
Matches 212; Conservative 127; Mismatches 243; Indels 464; Gaps 32;
916 OTLAENMEAGLSQQAQTLADYTABRLSSVLCNWFANIQPEGVSLHRDLDLYSYFLIDN 975
8 QTLKEARDAL-----VAHVIAQVPADLKE-----SIQTADDLYEYLLDIT 49

QY 976 QVSSAIKTRLAZAIAGIQLYINRAL-----NRIEPNARADVSTROFFTDW-TVNNRYSTW 1030
D 50 KISDLVTISPLSEAIAGSLQFIHRAIEGYDGTADSAKPYFADEQFLYNNWDSFNRHSTW 109
QY 1031 GGVSRLLVYPENYIDPTQRIQOTRMMDELLENISQSKLSRDTVEDAFKTYLTRFETVADL 1090
D 110 AKERLAFYAGDYIDPTLRNKTEIFTAFEGGISQSKLSSELVESKLRDYLISYDTLATL 169
QY 1091 KVSAYHDNVSNTGLTWFGVGTRENIPEYWRNVDISRMOAGELAAANAKWKITAV 1150
D 170 DYITACQCKDNKT---IFFIGRTQNAFYAFYWRKLTIV-TDGGKLPDQMSERAINAGI 225
QY 1151 N-PYKDAIRPVI PRERLHLIWEKEEVAKNGTDVETVDFRTLLKLAFLRHGDSWSAPWSY 1209
D 226 SEAYSGHVEFPWENKLIIRW-----FTI----- 249
QY 1210 DITTOVEAVTDKKPDTERLALAASFGQEDTLLVYVYTKGYSDFGSGNSKNQVAGMTI-Y 1268
D 250 -----SKEDKIDFVY-----KNIVWMSDY 269
QY 1269 GDGSKFKMENTALSRYSQLKNTFDIIHTQGNLVRKASYRPAQDFEVPASLNMGSALGDD 1328
D 270 SWASKKILELSFTDYNRVGAT-----GSSSPTEVASQYGS-----AQMNISD---DG 315
QY 1329 SLTWEN--GNIPOITSKYSSDNLAITLHNAAFVRYDGGSGNWRNKNQISAMKLTGVGDK 1386
D 316 TVLIFQAGGATP-----STGVTLCYD-SGNVIKN-----L 345
QY 1387 SOYGNAFIIANTVRHYGYSDLGGPITVYANKTKNYIASVOGHLNADYTRRLIITPVENN 1446
D 346 SSTGSANLSSKD-----YATTK--LRMCHGQSYN-----DNN 375
QY 1447 YYARLFEPPSPNTILMTVFTVGSNKTSDFKCKSYAVDGNNSQGFQIFSSYQSSGWLDD 1506
D 376 Y-----CNFTLSINTI-----BFTSY----- 391
QY 1507 TGINNTDIKITVMAGSKTHPTASDHIASLPANSFDPAMPYTFKPLEIDASSLAFTNNIAP 1566
D 392 -----GTFSSDGKQFTPPSGS----- 407
QY 1567 LDIVFETKAKDGRVLGKI KOTLSVKRVNYPEDILFLRETHSGAQYMQLVYRIRLNTLL 1626
D 408 -----AIDLHLFNY-----VDLNAALL 423
QY 1627 ASQLVSRANTGIDTILMTETORLPEPLGEGFFANFVLPKYDPAEHGDERWFKIHGNVG 1686
D 424 -----DISLDSLLNYDVO-----GQFG 440
QY 1687 GNTGRQPYSGMLSDTSETSMTLFVPVAGYMHGVRGVLGVGYQKITDYNTWESAFFYFD 1746
D 441 G----- 441
QY 1747 ETQKQFVLINDADHDSGMTQGGIVKNIKKYKGFNLVSIATGYSAPMD-FNSASALYYWEL 1805
D 442 -----SNPVDNFGPYGIYLWEI 459
QY 1806 FYTTPMCFORLLOKOFDEATQMINVYNPAGY-IVNGEI-----APWIKNCPLSETT 1859
D 460 FFHIPFLVTVMQTEORYEDATWYKIFRSAGYRDANGQLIMDGSKPRYVNVMPLODIT 519
QY 1860 SWNANPLDAIDPDVAQNDPMHYKIATFMRLDQLILRGDMAYRELTLDALNEAKWYVR 1919
D 520 AWDTTQPATDTPDVIMADPMHYKIAIFLHTDLLLIARGDSAYRQLERDITLVEAKWYIQ 579
QY 1920 TLELIGDEPEDYGSQOWAAPSLSGAA 1945
D 580 AQLLGPREDIHTTNTWNPNTLSKEA 605
RESULT 15
US-08-851-567B-34
; Sequence 34, Application US/08851567B

Qy	907	SNLP	SWDE	WQT	LAEN	MEAG	LS	TQ	930
		:	:	:	:	:	:	:	
		:	:	:	:	:	:	:	
Db	1078	ALLA	S---	----	VAQN	LSAA	ISNR	Q	1095

Search completed: February 16, 2006, 21:46:27
Job time : 69.1714 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:45:19 ; Search time 230.217 Seconds
(without alignments)
4606.314 Million cell updates/sec

Title: US-10-754-115-34
Perfect score: 13043
Sequence: 1 MYSTAVLNKISPTRDGQTM.....KALLESLSDIILHRYTIRS 2538

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Published Applications AA Main:
1: /cgm2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
2: /cgm2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
3: /cgm2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
4: /cgm2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
5: /cgm2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
6: /cgm2_6/ptodata/1/pubpaa/US11_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13043	100.0	2538	US-10-609-113-49	Sequence 49, Appl
2	13043	100.0	2538	US-10-753-901-20	Sequence 20, Appl
3	13043	100.0	2538	US-10-754-115-20	Sequence 20, Appl
4	13043	100.0	2538	US-10-754-115-34	Sequence 34, Appl
5	5496.5	42.1	2523	US-10-753-901-14	Sequence 14, Appl
6	5496.5	42.1	2523	US-10-754-115-14	Sequence 14, Appl
7	5043	38.7	2516	US-09-817-514A-2	Sequence 2, Appl
8	5043	38.7	2516	US-10-262-794A-47	Sequence 47, Appl
9	5043	38.7	2516	US-10-706-424-2	Sequence 2, Appl
10	5043	38.7	2516	US-10-703-280-4	Sequence 4, Appl
11	4986.5	38.2	2504	US-09-817-514A-8	Sequence 8, Appl
12	4986.5	38.2	2504	US-10-262-794A-12	Sequence 12, Appl
13	4986.5	38.2	2504	US-10-754-115-59	Sequence 59, Appl
14	4976	36.8	2534	US-11-020-848-11	Sequence 11, Appl
15	4802	36.8	2381	US-10-706-424-8	Sequence 8, Appl
16	4802	36.8	2381	US-10-754-115-63	Sequence 63, Appl
17	4574	35.1	2499	US-10-706-424-4	Sequence 4, Appl
18	4572	35.1	2499	US-10-754-115-62	Sequence 62, Appl
19	4433	34.0	2177	US-10-706-424-6	Sequence 6, Appl
20	3303	25.3	1849	US-10-262-794A-49	Sequence 49, Appl
21	3278.5	25.1	1844	US-10-262-794A-53	Sequence 53, Appl
22	2655	20.4	1273	US-11-020-848-6	Sequence 6, Appl
23	1643	12.6	1205	US-10-609-113-5	Sequence 5, Appl
24	1632	12.5	573	US-10-262-794A-55	Sequence 55, Appl
25	1615	12.4	579	US-10-262-794A-51	Sequence 51, Appl
26	1591	12.2	1565	US-10-262-794A-59	Sequence 59, Appl
27	1572.5	12.1	1208	US-10-609-113-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1

```

US-10-609-113-49
; Sequence 49, Application US/10609113
; Publication No. US20040110184A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Bevan, Scott
; APPLICANT: Zhu, Baolong
; TITLE OF INVENTION: Pesticidally Active Proteins and Polynucleotides Obtainable from
; TITLE OF INVENTION: Paenibacillus Species
; FILE REFERENCE: DAS-101XC2
; CURRENT APPLICATION NUMBER: US/10/609,113
; CURRENT FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: US 60/392,633
; PRIOR FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER: US 60/441,647
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 49
; LENGTH: 2538
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-609-113-49

```

```

Query Match      100.0%; Score 13043; DB 4; Length 2538;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYSTAVLNKISPTRDGQTM1AD1QYLSFSELKIFDDOLSWGEARHLYHETI1QKKN 60
DB 1 MYSTAVLNKISPTRDGQTM1AD1QYLSFSELKIFDDOLSWGEARHLYHETI1QKKN 60

QY 61 RLLEARIITRANPOLSGAIRLGIERSVRSYDENFMGARSSSFVKPGSVASMFSPAGYLT 120
DB 61 RLLEARIITRANPOLSGAIRLGIERSVRSYDENFMGARSSSFVKPGSVASMFSPAGYLT 120

QY 121 ELYREAKDLHFSSSAHYLDNRPPDLAD1TLGSNNMDTIST1TLNELLLEHITRTKGGD 180
DB 121 ELYREAKDLHFSSSAHYLDNRPPDLAD1TLGSNNMDTIST1TLNELLLEHITRTKGGD 180

QY 181 SDALMES1STYRQAITDTPYHQPETIRQIVIMTHDSTLSALSNPEVMQAGASL1A1LA 240
DB 181 SDALMES1STYRQAITDTPYHQPETIRQIVIMTHDSTLSALSNPEVMQAGASL1A1LA 240

QY 241 NISPELYNLTETETKADALFAQNFENITPENFASQSWIAKYGYGLSEVQKYLGL 300
DB 241 NISPELYNLTETETKADALFAQNFENITPENFASQSWIAKYGYGLSEVQKYLGL 300

```

QY 301 QNGYSDTSAYVDNI STGLVNNESKLEAYKI TRVKTDYDKNINYFDLMYEGNNOFFIR 360
DB 301 QNGYSDTSAYVDNI STGLVNNESKLEAYKI TRVKTDYDKNINYFDLMYEGNNOFFIR 360
QY 361 ANFKVSRFEGATLRKNAGPSGIVGSLGPIANTNFKSNYLSNISDSSEYKNGVKIYAYRY 420
DB 361 ANFKVSRFEGATLRKNAGPSGIVGSLGPIANTNFKSNYLSNISDSSEYKNGVKIYAYRY 420
QY 421 TSSTSATNQGGI FTTFESYPLTIIPALKNKAI RLCLTSGLSPNELQTIIVRSDNAQGIIND 480
DB 421 TSSTSATNQGGI FTTFESYPLTIIPALKNKAI RLCLTSGLSPNELQTIIVRSDNAQGIIND 480
QY 481 SVLTKVFYTLFPYSHRYALSFDQAOLVNGSVINQVADDDSVSHFNRLNTPPLKCKIPEAD 540
DB 481 SVLTKVFYTLFPYSHRYALSFDQAOLVNGSVINQVADDDSVSHFNRLNTPPLKCKIPEAD 540
QY 541 GNTVSI DPDEQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600
DB 541 GNTVSI DPDEQSTFARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600
QY 601 TLLARVHOLT VNELCMLYGLSPFNKKTASLSGSELPLRLVILWLYQVOWLTTEABITTEAI 660
DB 601 TLLARVHOLT VNELCMLYGLSPFNKKTASLSGSELPLRLVILWLYQVOWLTTEABITTEAI 660
QY 661 WLLCTPFSGNI SPEISNLNLANRPISSEDMAQSHNRELQAEIILAPFTIATLHLASPDMA 720
DB 661 WLLCTPFSGNI SPEISNLNLANRPISSEDMAQSHNRELQAEIILAPFTIATLHLASPDMA 720
QY 721 RYILLWTDNLRPGGLD IAGFMFTLVLKGSLNANETTQLVQFCHVMAQSLSVQTLRLSEAE 780
DB 721 RYILLWTDNLRPGGLD IAGFMFTLVLKGSLNANETTQLVQFCHVMAQSLSVQTLRLSEAE 780
QY 781 LSVLVI SGFV LGAKNOPAGQHNIDTLP SLYRFHQWINGLNGP GSDTLDMLRQOQTLADR 840
DB 781 LSVLVI SGFV LGAKNOPAGQHNIDTLP SLYRFHQWINGLNGP GSDTLDMLRQOQTLADR 840
QY 841 LASVWGLD I SMVTOAMYSAGVNOJCQWODINTVLQWIDVASALHTMPSVIRTLVNI RVT 900
DB 841 LASVWGLD I SMVTOAMYSAGVNOJCQWODINTVLQWIDVASALHTMPSVIRTLVNI RVT 900
QY 901 ALNKAESNLP SWDEWQTLAENNEAGLS TQQAQTLADYTAERLS SVLCNWF LANI QPEGVS 960
DB 901 ALNKAESNLP SWDEWQTLAENNEAGLS TQQAQTLADYTAERLS SVLCNWF LANI QPEGVS 960
QY 961 LHSRDDLYSYFLIDNQVSSAI KTRTLABAIAGIQLYINRALNR IEPNARADVSTRQFPTD 1020
DB 961 LHSRDDLYSYFLIDNQVSSAI KTRTLABAIAGIQLYINRALNR IEPNARADVSTRQFPTD 1020
QY 1021 WTVNRRYSTWGVSRLVYYPENYIDPTORI GQTRMDELLENISQS KLSRDTVEDAPKTY 1080
DB 1021 WTVNRRYSTWGVSRLVYYPENYIDPTORI GQTRMDELLENISQS KLSRDTVEDAPKTY 1080
QY 1081 LTRFETVADLKVSAYHDNVNSNTGLTW FVGQTRENLP EYYWRNV D I SRMQAGELAAANAW 1140
DB 1081 LTRFETVADLKVSAYHDNVNSNTGLTW FVGQTRENLP EYYWRNV D I SRMQAGELAAANAW 1140
QY 1141 KEWKIDTAVNPYKDAIRPVI FRERLHLI WVEKEBAVAKNGTDPVETYDRFTLKLAFLRHD 1200
DB 1141 KEWKIDTAVNPYKDAIRPVI FRERLHLI WVEKEBAVAKNGTDPVETYDRFTLKLAFLRHD 1200
QY 1201 GSWAPMSYDITTOVEA VTDKPPDTERLALAA SGFGQEDTLLV FVYKTKGKSYDFGGSNK 1260
DB 1201 GSWAPMSYDITTOVEA VTDKPPDTERLALAA SGFGQEDTLLV FVYKTKGKSYDFGGSNK 1260
QY 1261 NVAGMTIYGDGSFKKMENTALSRYSQ LKNFTFDI IHTQGNDLVRKASRFADQFEVPASLN 1320
DB 1261 NVAGMTIYGDGSFKKMENTALSRYSQ LKNFTFDI IHTQGNDLVRKASRFADQFEVPASLN 1320
QY 1321 MGSAGDDSLTWMENGNIPQITSKYSSDNLAITLHNAAFTRYDGS GNVIRNKQISAMKL 1380
DB 1321 MGSAGDDSLTWMENGNIPQITSKYSSDNLAITLHNAAFTRYDGS GNVIRNKQISAMKL 1380
QY 1381 TGVDGKSYGNAFI IANTV KHGGYSDLG GPIITVYNTKNYIASVQGHLMNADYTRRLIL 1440

DB 1381 TGVDGKSYGNAFI IANTV KHGGYSDLG GPIITVYNTKNYIASVQGHLMNADYTRRLIL 1440
QY 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGFQIFSSYQSS 1500
DB 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGSNKTSDFKKCSYAVDGNNSQGFQIFSSYQSS 1500
QY 1501 GMLDIDTGINNNTDIKITVMAGSKTHHTFASDH IASLPANSFDMPTFKPLEIDASSIAP 1560
DB 1501 GMLDIDTGINNNTDIKITVMAGSKTHHTFASDH IASLPANSFDMPTFKPLEIDASSIAP 1560
QY 1561 TNNIAPLDIVFETKAKDGRVLGKI KOTLSVKRVNYPN PEDILFLRTHSGAQYMLGVYRI 1620
DB 1561 TNNIAPLDIVFETKAKDGRVLGKI KOTLSVKRVNYPN PEDILFLRTHSGAQYMLGVYRI 1620
QY 1621 RLNTLLASQLVSRANTGIDTILTMETORLPEPP IGEFFANFVLPKYDPPAEHGDERWPKI 1680
DB 1621 RLNTLLASQLVSRANTGIDTILTMETORLPEPP IGEFFANFVLPKYDPPAEHGDERWPKI 1680
QY 1681 HIGNVGNTGROPYSGWLSDTSETSM TLFVPYAEGYMHGVR LGVGYOKI TYDNTWES 1740
DB 1681 HIGNVGNTGROPYSGWLSDTSETSM TLFVPYAEGYMHGVR LGVGYOKI TYDNTWES 1740
QY 1741 AFFYFDETKQOFVLINDADHD SGMTQOGI VKNIKKYKGF LNVSIATGY SAPMDFNSASAL 1800
DB 1741 AFFYFDETKQOFVLINDADHD SGMTQOGI VKNIKKYKGF LNVSIATGY SAPMDFNSASAL 1800
QY 1801 YYWELFYTTPMCMQORLQEKQFDEATOWIN VYVNPAGYI VNGBIAPWNCRPLEETTS 1860
DB 1801 YYWELFYTTPMCMQORLQEKQFDEATOWIN VYVNPAGYI VNGBIAPWNCRPLEETTS 1860
QY 1861 WNANPLDAIDPDVAQNDPMHYKIATFMRLLDQ IILRGDMAYRELTRDALNEAKWYVRT 1920
DB 1861 WNANPLDAIDPDVAQNDPMHYKIATFMRLLDQ IILRGDMAYRELTRDALNEAKWYVRT 1920
QY 1921 LELLGDEPEDYGSQWAAAPSLSGAASQTVQAA YQQDLTMLGRGVSXNLRANSVLGLFL 1980
DB 1921 LELLGDEPEDYGSQWAAAPSLSGAASQTVQAA YQQDLTMLGRGVSXNLRANSVLGLFL 1980
QY 1981 PEYNPALTDYQOTLURLRFLNLRHNL SIDGOPSLSAIYAEPTDPKALLTSMVQASQGGSAV 2040
DB 1981 PEYNPALTDYQOTLURLRFLNLRHNL SIDGOPSLSAIYAEPTDPKALLTSMVQASQGGSAV 2040
QY 2041 LPGTLSLYRFPVMLERTNRLVAQLTQFGTSL SSWAEHDDADELTLLLOQGMELATQSIR 2100
DB 2041 LPGTLSLYRFPVMLERTNRLVAQLTQFGTSL SSWAEHDDADELTLLLOQGMELATQSIR 2100
QY 2101 IQQRTVDEVDADI AVLAE SRRSAQNRL EKYQQLYDEDI NHGEQRAMSLDDAAAGSLAGQ 2160
DB 2101 IQQRTVDEVDADI AVLAE SRRSAQNRL EKYQQLYDEDI NHGEQRAMSLDDAAAGSLAGQ 2160
QY 2161 VLSTAEGVADLVPNVFGLACGSRWGAALRASAS VMSLSATASQY SAKISRSEAYRRRR 2220
DB 2161 VLSTAEGVADLVPNVFGLACGSRWGAALRASAS VMSLSATASQY SAKISRSEAYRRRR 2220
QY 2221 QEWEIFQRDNADGEVKQMDAQLESUKIR REAAQMOV EYQETQQAHTQAOLELQKRFITNKA 2280
DB 2221 QEWEIFQRDNADGEVKQMDAQLESUKIR REAAQMOV EYQETQQAHTQAOLELQKRFITNKA 2280
QY 2281 LYSWMRGKLSAIYYQFDDLTSQFC LMAQEALRRELTDNGVTFFIRGGAWNGTTAGIMAGET 2340
DB 2281 LYSWMRGKLSAIYYQFDDLTSQFC LMAQEALRRELTDNGVTFFIRGGAWNGTTAGIMAGET 2340
QY 2341 LLLNLAEWEMKWLBRDRALAEVTRTVSLAQPYQALSSDNFNLTBKLQFLBREGKGNVGAS 2400
DB 2341 LLLNLAEWEMKWLBRDRALAEVTRTVSLAQPYQALSSDNFNLTBKLQFLBREGKGNVGAS 2400
QY 2401 GNEUKLSNRQITEASVRISDLKIFSDY PEGSLGNTQLKQVSVTLPALVGPYEDIRAVLNYG 2460
DB 2401 GNEUKLSNRQITEASVRISDLKIFSDY PEGSLGNTQLKQVSVTLPALVGPYEDIRAVLNYG 2460
QY 2461 GSIYVMPGCSAIALSHGVNDSGQFM LDFNDSRYLPFEGISVNDSGSLTSPFDPATDROKA 2520

Db 2461 GSIVMPGCSAIALSHGVNDSCQFMLDFNDSRYLPFEGISVNDSSGLTSLSPFDATDROKA 2520

Qy 2521 LLESLSDIILHIRTIRS 2538

Db 2521 LLESLSDIILHIRTIRS 2538

RESULT 2

US-10-753-901-20

; Sequence 20, Application US/10753901

; Publication No. US200401941641

; GENERAL INFORMATION:

; APPLICANT: Bintrim, Scott

; APPLICANT: Mitchell, Jon

; APPLICANT: Larrinua, Ignacio

; APPLICANT: Apel-Birkhold, Patricia

; APPLICANT: Schafer, Barry

; APPLICANT: Bevan, Scott

; APPLICANT: Young, Scott

; APPLICANT: Guo, Lining

; TITLE OF INVENTION: Xenorhabdus TC Proteins and Genes for Pest Control

; FILE REFERENCE: DAS-105X

; CURRENT APPLICATION NUMBER: US/10/753,901

; CURRENT FILING DATE: 2004-01-07

; PRIOR APPLICATION NUMBER: US 60/441,717

; PRIOR FILING DATE: 2003-01-21

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn version 3.2

; SEQ ID NO 20

; LENGTH: 2538

; TYPE: PRT

; ORGANISM: Xenorhabdus nematophilus

US-10-753-901-20

Query Match 100.0%; Score 13043; DB 4; Length 2538;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MYSTAVLANKISPTRDGQTMTLADLQYLSFSELRIKIFDDQLSWGAEARHLXHTTIOKQNN 60

Db 1 MYSTAVLANKISPTRDGQTMTLADLQYLSFSELRIKIFDDQLSWGAEARHLXHTTIOKQNN 60

Qy 61 RLLEARIPTRANPOLSGAIRLGIERSVRSYDEMFGARSSSFVKPGSVASMFSPAGYLT 120

Db 61 RLLEARIPTRANPOLSGAIRLGIERSVRSYDEMFGARSSSFVKPGSVASMFSPAGYLT 120

Qy 121 ELYREAKDLHFSSAYHLDNRRLADLTLQSNDTETISLTLSNELLLLEHITRKTGCD 180

Db 121 ELYREAKDLHFSSAYHLDNRRLADLTLQSNDTETISLTLSNELLLLEHITRKTGCD 180

Qy 181 SDALMESLSTYRQADTPHYOPIETIROVIMTHDSTLSALSRNPVWMOAGASLLAILA 240

Db 181 SDALMESLSTYRQADTPHYOPIETIROVIMTHDSTLSALSRNPVWMOAGASLLAILA 240

Qy 241 NISPELYNLTETETKRNADALFAQNFSENITPENFASQSWIAKYIGLESEVQKYLGM 300

Db 241 NISPELYNLTETETKRNADALFAQNFSENITPENFASQSWIAKYIGLESEVQKYLGM 300

Qy 301 QNGYSDSTSAVDNITSTGLVNNESKLEAYKIIRVKTDDYDKNINFDLMYEGNNOFFIR 360

Db 301 QNGYSDSTSAVDNITSTGLVNNESKLEAYKIIRVKTDDYDKNINFDLMYEGNNOFFIR 360

Qy 361 ANFKVSRFEGATLRKNAGPSGLVSGPLANTNFKNYLSNISDSYKNGVKIYAYRY 420

Db 361 ANFKVSRFEGATLRKNAGPSGLVSGPLANTNFKNYLSNISDSYKNGVKIYAYRY 420

Qy 421 TSSTSATNQGGGIFTFESYPLTIFALKNKAIRCLCTSLSPNELQTTIVRSDNAQGIIND 480

Db 421 TSSTSATNQGGGIFTFESYPLTIFALKNKAIRCLCTSLSPNELQTTIVRSDNAQGIIND 480

Qy 481 SVLTAKVFTLTFYSHRYALSFDDAQVLNGSVINQYADDSVSHFNRLFNTPLPKGKIFEAD 540

Db 481 SVLTAKVFTLTFYSHRYALSFDDAQVLNGSVINQYADDSVSHFNRLFNTPLPKGKIFEAD 540

Qy 541 GNTVSIDPDERQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL 600

Db 541 GNTVSIDPDERQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL 600

Qy 601 TLLARVHQLTVNELCMLYGLSPFNKTKTASLSSGELPRLVWLYQVOTWLTAEATTEAI 660

Db 601 TLLARVHQLTVNELCMLYGLSPFNKTKTASLSSGELPRLVWLYQVOTWLTAEATTEAI 660

Qy 661 WLLCTPFESGNI SPEISNLNLRPSISSEDMAQSHNRELOAEILAPFTAATLHLASPDWA 720

Db 661 WLLCTPFESGNI SPEISNLNLRPSISSEDMAQSHNRELOAEILAPFTAATLHLASPDWA 720

Qy 721 RYILLWTDNLRPGGLDIAGFMTLVLKESLNANETTQLVQFCHVMAQLSLSVQTLRLSEAE 780

Db 721 RYILLWTDNLRPGGLDIAGFMTLVLKESLNANETTQLVQFCHVMAQLSLSVQTLRLSEAE 780

Qy 781 LSVLVISGFVILGAKNQAGOHNDITLFSLYRFHQWINGLGNPGSDTTLMLRQOQTLTADR 840

Db 781 LSVLVISGFVILGAKNQAGOHNDITLFSLYRFHQWINGLGNPGSDTTLMLRQOQTLTADR 840

Qy 841 LASVMGLDISMVTQAMVSAGVNLQOCQWODINTVLQWIDVASALHTMPSVIRTLVNIYYT 900

Db 841 LASVMGLDISMVTQAMVSAGVNLQOCQWODINTVLQWIDVASALHTMPSVIRTLVNIYYT 900

Qy 901 ALNKAESNLPSEWDEWQTLAENNEAGLSLQOQTLADYTAERLSSVLCNWFNLANIQPEGVS 960

Db 901 ALNKAESNLPSEWDEWQTLAENNEAGLSLQOQTLADYTAERLSSVLCNWFNLANIQPEGVS 960

Qy 961 LHSRDDLYSYFLIDNQVSSAIKTRTLABAIAGIQLYINRALNRIEPNARADVSTQFFTD 1020

Db 961 LHSRDDLYSYFLIDNQVSSAIKTRTLABAIAGIQLYINRALNRIEPNARADVSTQFFTD 1020

Qy 1021 WTVNNRSTYMGVSRVLYPENYIDPTORIGOTRMDDELLENISOSKLSRDTVEDAFKTY 1080

Db 1021 WTVNNRSTYMGVSRVLYPENYIDPTORIGOTRMDDELLENISOSKLSRDTVEDAFKTY 1080

Qy 1081 LTRFETVADLVKVSAYHDNVNSNTGLTWFGVQOTRENLPYVYWRNVYDLSRMQAGELAANAW 1140

Db 1081 LTRFETVADLVKVSAYHDNVNSNTGLTWFGVQOTRENLPYVYWRNVYDLSRMQAGELAANAW 1140

Qy 1141 KEWKIDITAVNPYKDAIRPVI FRERLHLI WYEKEEVAKNGTDPVETYDRFTLKLAFLRH 1200

Db 1141 KEWKIDITAVNPYKDAIRPVI FRERLHLI WYEKEEVAKNGTDPVETYDRFTLKLAFLRH 1200

Qy 1201 GSWAPWSYDITTOVEAVTDKPPDTERLALAASGFGEDTLLVFYKTKGKSYSDFGGSK 1260

Db 1201 GSWAPWSYDITTOVEAVTDKPPDTERLALAASGFGEDTLLVFYKTKGKSYSDFGGSK 1260

Qy 1261 NVAGMTIYGDGSKFKMENTALSRYSQLKNTFDI IHTQGNDLVRKASYPFAQDFEVPASLN 1320

Db 1261 NVAGMTIYGDGSKFKMENTALSRYSQLKNTFDI IHTQGNDLVRKASYPFAQDFEVPASLN 1320

Qy 1321 MGSAGDSDLTVMENGINPQITSKYSSDNLAITLHNAAFTRYDGSNGVIRNKQISAMKL 1380

Db 1321 MGSAGDSDLTVMENGINPQITSKYSSDNLAITLHNAAFTRYDGSNGVIRNKQISAMKL 1380

Qy 1381 TGVGKSOYGNAFIITANTVKHYGYSGLGGPTTVNKTKNYIASVQGHLMNADYTRRLIL 1440

Db 1381 TGVGKSOYGNAFIITANTVKHYGYSGLGGPTTVNKTKNYIASVQGHLMNADYTRRLIL 1440

Qy 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGSNKTSDFKCSYAVDGNNSQGFQIFSSYQSS 1500

Db 1441 TPVENNYARLFEPPFPSPNTILNTVFTVGSNKTSDFKCSYAVDGNNSQGFQIFSSYQSS 1500

Qy 1501 GWLDDITGINNTDITKITVMAGSKTHFTTASDHIAISLPANSDAMPYTKPLEIDASSLAF 1560

Db 1501 GWLDDITGINNTDITKITVMAGSKTHFTTASDHIAISLPANSDAMPYTKPLEIDASSLAF 1560

Qy 1561 TNNIAPLDIVFETAKOGRVLGKIQTLSVKRVNVPEDILFLRTHSGAQMVLGVYRI 1620

Db 1561 TNNIAPLDIVFETAKOGRVLGKIQTLSVKRVNVPEDILFLRTHSGAQMVLGVYRI 1620

```
Qy 1621 RLNTLLASQVSRANTGIDTILTMETORLPBPPLGEGFFANFVLPKYPDPAEGDERWFKI 1680
Db 1621 RLNTLLASQVSRANTGIDTILTMETORLPBPPLGEGFFANFVLPKYPDPAEGDERWFKI 1680
Qy 1681 HIGNVGGNTGRQPYSGMLSDTSETMTLPVPAEAGYMHGVRGLGVGYKITDYNTWES 1740
Db 1681 HIGNVGGNTGRQPYSGMLSDTSETMTLPVPAEAGYMHGVRGLGVGYKITDYNTWES 1740
Qy 1741 AFFYFDETKQOVLINDADHDGSGTQOQIVKNIKKYKGFNLVSIATGYSAPMDNFSASAL 1800
Db 1741 AFFYFDETKQOVLINDADHDGSGTQOQIVKNIKKYKGFNLVSIATGYSAPMDNFSASAL 1800
Qy 1801 YYWELFYTPMWCQRLLQKQFQDEATQWIVNVPAGYI VNGHIAPIWNCRLPEETTS 1860
Db 1801 YYWELFYTPMWCQRLLQKQFQDEATQWIVNVPAGYI VNGHIAPIWNCRLPEETTS 1860
Qy 1861 WNANPLDAIDPDAVAQNPDPMHYKIATFMRLLDQILRGDMAYRELTRDALNEAKWYVRT 1920
Db 1861 WNANPLDAIDPDAVAQNPDPMHYKIATFMRLLDQILRGDMAYRELTRDALNEAKWYVRT 1920
Qy 1921 LELLGDEPEDYGSQWAAAPSLSGAASQTVOAAYQODLTMLGRGGVSKNLRITANSIVGLFL 1980
Db 1921 LELLGDEPEDYGSQWAAAPSLSGAASQTVOAAYQODLTMLGRGGVSKNLRITANSIVGLFL 1980
Qy 1981 PEYNPALTDYQWTLRLRLFNLRHNLSDIGQPLSLAI VAEPDTPKALLTSMVQASQGGSAV 2040
Db 1981 PEYNPALTDYQWTLRLRLFNLRHNLSDIGQPLSLAI VAEPDTPKALLTSMVQASQGGSAV 2040
Qy 2041 LPGTLSLYRFPVMLERTNLVAQLTQFGTSLLSMAEHDDADDELTTLLQQGMELATQSIR 2100
Db 2041 LPGTLSLYRFPVMLERTNLVAQLTQFGTSLLSMAEHDDADDELTTLLQQGMELATQSIR 2100
Qy 2101 IQQRTVDEVDADIIVLAESRRSAQNRLKYQOQYDEIDINHGEOQAMSLDDAAQCSLAGQ 2160
Db 2101 IQQRTVDEVDADIIVLAESRRSAQNRLKYQOQYDEIDINHGEOQAMSLDDAAQCSLAGQ 2160
Qy 2161 VLSIABGVADLVPNVFGLAGCGSRWGAALRASAVMSLSATASOVSADKISRSAYRRRR 2220
Db 2161 VLSIABGVADLVPNVFGLAGCGSRWGAALRASAVMSLSATASOVSADKISRSAYRRRR 2220
Qy 2221 QEWEIFQDNADGEVKQMDAQLESILKIRREAAQMVQEYQEQTAQHTQAQLELLQKFTNKA 2280
Db 2221 QEWEIFQDNADGEVKQMDAQLESILKIRREAAQMVQEYQEQTAQHTQAQLELLQKFTNKA 2280
Qy 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEARRELTNGVTFIRGGAWNGTTAGIMAGET 2340
Db 2281 LYSWMRGKLSAIYYQFFDLTQSFCLMAQEARRELTNGVTFIRGGAWNGTTAGIMAGET 2340
Qy 2341 LLLNLAEMEKVWLERDERALEBVTITVSLAQFYQALSSDNFNLTEKLTOFLREGKNGVAS 2400
Db 2341 LLLNLAEMEKVWLERDERALEBVTITVSLAQFYQALSSDNFNLTEKLTOFLREGKNGVAS 2400
Qy 2401 GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLTPALVGPYEDIRAVLNYG 2460
Db 2401 GNELKLSNRQTEASVRLSDLKIFSDYPESLGNTRQLKQVSVTLTPALVGPYEDIRAVLNYG 2460
Qy 2461 GSIWMPRGCSAIALSHGWNDSQGMFLDFNDNRYLPFPEGISVNDGSGSLTSPDPTDRQKA 2520
Db 2461 GSIWMPRGCSAIALSHGWNDSQGMFLDFNDNRYLPFPEGISVNDGSGSLTSPDPTDRQKA 2520
Qy 2521 LLESLSDIILHIRTIRS 2538
Db 2521 LLESLSDIILHIRTIRS 2538
```

RESULT 3

```
US-10-754-115-20
; Sequence 20, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
```

```
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, BaoLong
; APPLICANT: Mezlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XCI
; CURRENT APPLICATION NUMBER: US/10754,115
; PRIOR FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 2538
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-754-115-20
```

```
Query Match 100.0%; Score 13043; DB 4; Length 2538;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy 1 MYSTAVLLNKISPRDGOQWTMLADLOVLSRSELKIFDDQLSWGFEARHLVHETIEOKKN 60
Db 1 MYSTAVLLNKISPRDGOQWTMLADLOVLSRSELKIFDDQLSWGFEARHLVHETIEOKKN 60
Qy 61 RLLEARLFTFRANPOLSGAIRLGIERSVRSYDMDFGARSSSFVKPGSVASMFSPAGYLT 120
Db 61 RLLEARLFTFRANPOLSGAIRLGIERSVRSYDMDFGARSSSFVKPGSVASMFSPAGYLT 120
Qy 121 ELYREAKDLHFSAYHLNRRPDLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180
Db 121 ELYREAKDLHFSAYHLNRRPDLADLTLSQSNMDEISTLTLSNELLLEHITRKTGGD 180
Qy 181 SDALMESLSTYRQAIPTYHQPYETIRQVIMTHDSTLSALSRNPEVWGQAGSALLAILA 240
Db 181 SDALMESLSTYRQAIPTYHQPYETIRQVIMTHDSTLSALSRNPEVWGQAGSALLAILA 240
Qy 241 NISPELXNLTETETKXNADALFAQNFSEMITPENFASQSWIAKYVGLSEVQKYLGM 300
Db 241 NISPELXNLTETETKXNADALFAQNFSEMITPENFASQSWIAKYVGLSEVQKYLGM 300
Qy 301 QNGVSDSTSAVVDNISTGLVWNESKLEAYKITRKTDDYDKNINYPDLMEYGNQPFIR 360
Db 301 QNGVSDSTSAVVDNISTGLVWNESKLEAYKITRKTDDYDKNINYPDLMEYGNQPFIR 360
Qy 361 ANFKVSRFGATLERNKAGPSGIYVGLSGPLIANTNFKSNVLSNDSSEYKNGVKIYVRY 420
Db 361 ANFKVSRFGATLERNKAGPSGIYVGLSGPLIANTNFKSNVLSNDSSEYKNGVKIYVRY 420
Qy 421 TSSTSATNQGGIIFTFESYPLTIPALKLNKAIKRLTSLGSPNELQITVRSDNAQGIIND 480
Db 421 TSSTSATNQGGIIFTFESYPLTIPALKLNKAIKRLTSLGSPNELQITVRSDNAQGIIND 480
Qy 481 SVLTQVFTYTLFYSHRYALSFDDAQVLNGSVINQVADDDSVSHFNRLFTNTPPLKGIPEAD 540
Db 481 SVLTQVFTYTLFYSHRYALSFDDAQVLNGSVINQVADDDSVSHFNRLFTNTPPLKGIPEAD 540
Qy 541 GNTVSDIDPDEEQSTFARSALMRGIVGNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600
Db 541 GNTVSDIDPDEEQSTFARSALMRGIVGNSGELYQLGKLAGVLDAQNTITLSVFVSSLYRL 600
Qy 601 TLLARVHQLTVNELCMLYGLSPFNKGT TASLSSGELPRLVILWYQVOWLTAEAITTEAI 660
Db 601 TLLARVHQLTVNELCMLYGLSPFNKGT TASLSSGELPRLVILWYQVOWLTAEAITTEAI 660
Qy 661 WLLCTPFGSGNISPEISNLLNLPSPISSEDMAQSHNRELOAEIILAPTAATLHLASPDMA 720
Db 661 WLLCTPFGSGNISPEISNLLNLPSPISSEDMAQSHNRELOAEIILAPTAATLHLASPDMA 720
```

QY 721 RYLLWTNLRPGGLDIAGFMTLVKESLNANETTQLVQFCHVMAQLSLSVQTLRLSEAE 780
DB 721 RYLLWTNLRPGGLDIAGFMTLVKESLNANETTQLVQFCHVMAQLSLSVQTLRLSEAE 780
QY 781 LSVLVISGFVILGAKNQAGOHNDITLPSLRFHOWINGLNGPSDDTLDMRLQOQTLTADR 840
DB 781 LSVLVISGFVILGAKNQAGOHNDITLPSLRFHOWINGLNGPSDDTLDMRLQOQTLTADR 840
QY 841 LASVGLDLSMVTQAMWSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNRVYT 900
DB 841 LASVGLDLSMVTQAMWSAGVNLQOCQODINTVLQWIDVASALHTMPSVIRTLVNRVYT 900
QY 901 ALNKAESNLPSWDEWQTLAENMEAGLSLTOAQTLADYTAERLSSVLCNFWLANIOPEGVS 960
DB 901 ALNKAESNLPSWDEWQTLAENMEAGLSLTOAQTLADYTAERLSSVLCNFWLANIOPEGVS 960
QY 961 LHSRDDLYSYFLIDNQVSSAIKTRTLAELAIAGIQLYINRALNRPENARADVSTROQFTD 1020
DB 961 LHSRDDLYSYFLIDNQVSSAIKTRTLAELAIAGIQLYINRALNRPENARADVSTROQFTD 1020
QY 1021 WTVNRRYSTGWSRLVYPENYIDPTQIRIGOTRMDELLENISQSLSRDTVEDAPKTY 1080
DB 1021 WTVNRRYSTGWSRLVYPENYIDPTQIRIGOTRMDELLENISQSLSRDTVEDAPKTY 1080
QY 1081 LTRFETVADLVKWSAYHDNVNSNTGLTWFGQGTRENLPEYTYRNVDIISRMQAGELAAANAW 1140
DB 1081 LTRFETVADLVKWSAYHDNVNSNTGLTWFGQGTRENLPEYTYRNVDIISRMQAGELAAANAW 1140
QY 1141 KEWKIDTAVNPYKDAIRPVIIFRERLHLI WYKEEVAKNKGDPTVETDRFTLKLAFLRHD 1200
DB 1141 KEWKIDTAVNPYKDAIRPVIIFRERLHLI WYKEEVAKNKGDPTVETDRFTLKLAFLRHD 1200
QY 1201 GSWAPWSYDITTOVEAVTDKPPDTERLALASGFGQEDTLLVFKYTKGKSYDFGGSNK 1260
DB 1201 GSWAPWSYDITTOVEAVTDKPPDTERLALASGFGQEDTLLVFKYTKGKSYDFGGSNK 1260
QY 1261 NVAGMTIYDGSFKKMENTALSRYSQLKNTFDIIHTQGNDLVRKASVFAQDFEVPASLN 1320
DB 1261 NVAGMTIYDGSFKKMENTALSRYSQLKNTFDIIHTQGNDLVRKASVFAQDFEVPASLN 1320
QY 1321 MSAIGDLSLTMENGNIPQITSYSSDNLAITLHNAATFVYDGSNVIRNKQISAMKL 1380
DB 1321 MSAIGDLSLTMENGNIPQITSYSSDNLAITLHNAATFVYDGSNVIRNKQISAMKL 1380
QY 1381 TGVGKSYGNAFIANTVKHYGYSGLGGPITVYNTKNTKNIASVQGHLMADYTRRLIL 1440
DB 1381 TGVGKSYGNAFIANTVKHYGYSGLGGPITVYNTKNTKNIASVQGHLMADYTRRLIL 1440
QY 1441 TPVENNYARLFEPPFSPNTILNTVFTVGSNKTSDFKCKSVAVDGNNSQGFQIFSSYQSS 1500
DB 1441 TPVENNYARLFEPPFSPNTILNTVFTVGSNKTSDFKCKSVAVDGNNSQGFQIFSSYQSS 1500
QY 1501 GWLIDIDGINNTDITKITWAGSKTHFTASDHIALSPANSFDMPTFKPLIEIDASSLAF 1560
DB 1501 GWLIDIDGINNTDITKITWAGSKTHFTASDHIALSPANSFDMPTFKPLIEIDASSLAF 1560
QY 1561 TNNIAPLDIVPETKAKOGRVLGKIKOTLSVKRVNPNEDILFLRETHSGAOYMWLVYRI 1620
DB 1561 TNNIAPLDIVPETKAKOGRVLGKIKOTLSVKRVNPNEDILFLRETHSGAOYMWLVYRI 1620
QY 1621 RLNTLLASQLVSRANTGIDITLTMTORLPBPPLGEGFFANFVLPKYDPAEHBGDERWPKI 1680
DB 1621 RLNTLLASQLVSRANTGIDITLTMTORLPBPPLGEGFFANFVLPKYDPAEHBGDERWPKI 1680
QY 1681 HIGNVGGTGRQPYSGMLSDTSETMTLFPYAEGYTHMEGVRLGVGYKITVDNTWES 1740
DB 1681 HIGNVGGTGRQPYSGMLSDTSETMTLFPYAEGYTHMEGVRLGVGYKITVDNTWES 1740
QY 1741 AFFYFDETQOQFVLINDADHDSGMTQOQIGVNIKKYKGLNVSATGYSAPMDFNASAL 1800
DB 1741 AFFYFDETQOQFVLINDADHDSGMTQOQIGVNIKKYKGLNVSATGYSAPMDFNASAL 1800

QY 1801 YWELFYFTYPMCFORLLQEKOFDEATQWINTVYNPAGYIVNGEITAPWIMNCRPLEETTS 1860
DB 1801 YWELFYFTYPMCFORLLQEKOFDEATQWINTVYNPAGYIVNGEITAPWIMNCRPLEETTS 1860
QY 1861 WNAWPLDAIDPDAVAQNDPMHYKIATFMRLLDQDLILRGDMAYRELTRDALNAEKWYVRT 1920
DB 1861 WNAWPLDAIDPDAVAQNDPMHYKIATFMRLLDQDLILRGDMAYRELTRDALNAEKWYVRT 1920
QY 1921 LELLAGDEPEDYGSQQAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSILVGLFL 1980
DB 1921 LELLAGDEPEDYGSQQAAPSLSGAASQTVQAAAYQODLTMLGRGGVSKNLRANSILVGLFL 1980
QY 1981 PEYNPALTDYQWTLRLRLFNLRHNLSDIGQPLSLAIYAEPDTPKALLTSMVQASOGGSV 2040
DB 1981 PEYNPALTDYQWTLRLRLFNLRHNLSDIGQPLSLAIYAEPDTPKALLTSMVQASOGGSV 2040
QY 2041 LPGTLISLYRFPVLMERTNLVAOLTQFGTSLLSWAHDDADDELTTLLQQQGMELATQISIR 2100
DB 2041 LPGTLISLYRFPVLMERTNLVAOLTQFGTSLLSWAHDDADDELTTLLQQQGMELATQISIR 2100
QY 2101 IQORTVDEVDADIIVLAESRBSAQNRLEKYQQLYDEDEDINHGEQRAMSLDDAAAGQSLAQ 2160
DB 2101 IQORTVDEVDADIIVLAESRBSAQNRLEKYQQLYDEDEDINHGEQRAMSLDDAAAGQSLAQ 2160
QY 2161 VLSIAEGVADIVPNVFGGLACGSRWGAALRASAVMSLSATASQYSADKISRSEAYRRRR 2220
DB 2161 VLSIAEGVADIVPNVFGGLACGSRWGAALRASAVMSLSATASQYSADKISRSEAYRRRR 2220
QY 2221 QEWEIFQRDNADGEVQKQMDAQLSLEKIRREAAQMOVEYOETQOANTQAOLELLQKFTNKA 2280
DB 2221 QEWEIFQRDNADGEVQKQMDAQLSLEKIRREAAQMOVEYOETQOANTQAOLELLQKFTNKA 2280
QY 2281 LYSWNRGKLSAIYYQFFDLTQSFCLMAQEARLRELTONGVTFIRGAWNGTTAGLMAGET 2340
DB 2281 LYSWNRGKLSAIYYQFFDLTQSFCLMAQEARLRELTONGVTFIRGAWNGTTAGLMAGET 2340
QY 2341 LLLNLAEMEKVWLBRRDERALEVTRTVSLAQFYQALSSDNFNLTETKLTQFLREGKGNVGAS 2400
DB 2341 LLLNLAEMEKVWLBRRDERALEVTRTVSLAQFYQALSSDNFNLTETKLTQFLREGKGNVGAS 2400
QY 2401 GNEUKLSNRQIEASVRLSDLKI PSDYPESLGNTRQLKQSVTLTPALVGPYEDIRAVLNG 2460
DB 2401 GNEUKLSNRQIEASVRLSDLKI PSDYPESLGNTRQLKQSVTLTPALVGPYEDIRAVLNG 2460
QY 2461 GSIWVRGCSAIALSHGVNDSGQFMDFNDSDRYLPFEGISVNDSGSLTSPFPDATDROKA 2520
DB 2461 GSIWVRGCSAIALSHGVNDSGQFMDFNDSDRYLPFEGISVNDSGSLTSPFPDATDROKA 2520
QY 2521 LLESLSDIILHIRVTIRS 2538
DB 2521 LLESLSDIILHIRVTIRS 2538

RESULT 4

US-10-754-115-34
; Sequence 34, Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Bintirim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XCI
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07

; PRIOR APPLICATION NUMBER: US 60/441,723									
; PRIOR FILING DATE: 2003-01-21									
; NUMBER OF SEQ ID NOS: 64									
; SOFTWARE: Patent in version 3.2									
; SEQ ID NO 34									
; LENGTH: 2538									
; TYPE: PRF									
; ORGANISM: Xenorhabdus nematophilus									
US-10-754-115-34									
Query Match 100.0%; Score 13043; DB 4; Length 2538;									
Beet Local Similarity 100.0%; Pred. No. 0;									
Matches 2538; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	MYSTAVILNKISPTRDGQWTMLADQLYLSFELARKIFDDQLSWGAEARHLVHETIEQKNN	60						
Db	1	MYSTAVILNKISPTRDGQWTMLADQLYLSFELARKIFDDQLSWGAEARHLVHETIEQKNN	60						
Qy	61	RLLEARIPTTRANPOLSGAIRLGIERDSVRSYDEMFGARSSSVKPGSVASMPSPAGYLT	120						
Db	61	RLLEARIPTTRANPOLSGAIRLGIERDSVRSYDEMFGARSSSVKPGSVASMPSPAGYLT	120						
Qy	121	ELYREAKDLHPSSAYHLDNRRLADLTLSQSNMDEISTLSNELLELLEHITRKTTGGD	180						
Db	121	ELYREAKDLHPSSAYHLDNRRLADLTLSQSNMDEISTLSNELLELLEHITRKTTGGD	180						
Qy	181	SDALMESLSTYRQAITDTPHYOPYETIRQVIMTHDSTLSALSRNPVEMQAGASILLAILA	240						
Db	181	SDALMESLSTYRQAITDTPHYOPYETIRQVIMTHDSTLSALSRNPVEMQAGASILLAILA	240						
Qy	241	NISPELYNILTEETETKKNADALFAQNFSENITPENFASQSWIAKYIGLESEVOKYLGML	300						
Db	241	NISPELYNILTEETETKKNADALFAQNFSENITPENFASQSWIAKYIGLESEVOKYLGML	300						
Qy	301	QNGYSDSTSAYVDNISGLVNNESKLEAYKITRVKTDYDKNINYPDLMEYEGNNQFFIR	360						
Db	301	QNGYSDSTSAYVDNISGLVNNESKLEAYKITRVKTDYDKNINYPDLMEYEGNNQFFIR	360						
Qy	361	ANPKVSRFEGATLRKNAGPSGIVGSLGPLIANTNFKSNYLSNISDSSEYKNGVKIYAYRY	420						
Db	361	ANPKVSRFEGATLRKNAGPSGIVGSLGPLIANTNFKSNYLSNISDSSEYKNGVKIYAYRY	420						
Qy	421	TSSTSATNQGGGIFTFESYPLTIPALKNKAI RLCLTSGLSPNELQTTIVRSDNAGGIIND	480						
Db	421	TSSTSATNQGGGIFTFESYPLTIPALKNKAI RLCLTSGLSPNELQTTIVRSDNAGGIIND	480						
Qy	481	SVLTAKVFYTLFYSHRYALSPDQAQVLNGSVINQYADDDSVSHFNRLFNTPPLKGKIFEAD	540						
Db	481	SVLTAKVFYTLFYSHRYALSPDQAQVLNGSVINQYADDDSVSHFNRLFNTPPLKGKIFEAD	540						
Qy	541	GNTVSIIDPDEQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL	600						
Db	541	GNTVSIIDPDEQSTFARSALMRGLGVNSGELYQLGKLAGVLDQAQNTITLSVFVSSLYRL	600						
Qy	601	TLARVHQLTYNELCMLGSLFPNGKTTASLSSELPRLVTLVQVOWLFEASITTEAI	660						
Db	601	TLARVHQLTYNELCMLGSLFPNGKTTASLSSELPRLVTLVQVOWLFEASITTEAI	660						
Qy	661	WLLCTPEFSGNISPEISNLLNLRPSISEDMAQSHNRELQAEILAPPTAATLHLASPDMA	720						
Db	661	WLLCTPEFSGNISPEISNLLNLRPSISEDMAQSHNRELQAEILAPPTAATLHLASPDMA	720						
Qy	721	RYILLWTNLRPPGLDIAQFMVLVLESINANETTLQVQFCHVAQLSLSVQTLRLSEAE	780						
Db	721	RYILLWTNLRPPGLDIAQFMVLVLESINANETTLQVQFCHVAQLSLSVQTLRLSEAE	780						
Qy	781	LSVLVISGFAVLGAQNPAGOHNTDITLSLYRFHOWINGLNGPSSDITLMDLRQOQTLTADR	840						
Db	781	LSVLVISGFAVLGAQNPAGOHNTDITLSLYRFHOWINGLNGPSSDITLMDLRQOQTLTADR	840						
Qy	841	LASVMGLDISMVTQAMVSAGVNLQCCWQDINTVLQWIDVASALHTMPSVIRTLVNIYYVT	900						
Db	841	LASVMGLDISMVTQAMVSAGVNLQCCWQDINTVLQWIDVASALHTMPSVIRTLVNIYYVT	900						

Qy	901	ALNKAESNLPSWDEWQTLAENMEAGLSTQAOQTLADYTAERLSSVLCNWFLANIQPEGVS	960						
Db	901	ALNKAESNLPSWDEWQTLAENMEAGLSTQAOQTLADYTAERLSSVLCNWFLANIQPEGVS	960						
Qy	961	LHSRDDLYSYFLIDNQVSSAIKTRRLAEAIAGIQLYINRALNRIEPNARADVSTROFFTD	1020						
Db	961	LHSRDDLYSYFLIDNQVSSAIKTRRLAEAIAGIQLYINRALNRIEPNARADVSTROFFTD	1020						
Qy	1021	WTVNNRYSTWGVSRVLYYPENYIDPTQRIGQTRMDELLENISQSKLSDRTVEDAFKTY	1080						
Db	1021	WTVNNRYSTWGVSRVLYYPENYIDPTQRIGQTRMDELLENISQSKLSDRTVEDAFKTY	1080						
Qy	1081	LTRPETVADLKVVSAVYHDNVNSNTGLTFVQGTRENLPEYYWRNVDISRWQAGELANAW	1140						
Db	1081	LTRPETVADLKVVSAVYHDNVNSNTGLTFVQGTRENLPEYYWRNVDISRWQAGELANAW	1140						
Qy	1141	KEWTKIDTAVNPYKDAIRPVI FRERLHLI WVEKEEVAKNGTDPVETYDRFTLKLAFLRHD	1200						
Db	1141	KEWTKIDTAVNPYKDAIRPVI FRERLHLI WVEKEEVAKNGTDPVETYDRFTLKLAFLRHD	1200						
Qy	1201	GSWSAPWSYDITTOVEAVTDKKPDTERLALAAAGFQGEDTLLVPVYKTGKSYSPFGGSNK	1260						
Db	1201	GSWSAPWSYDITTOVEAVTDKKPDTERLALAAAGFQGEDTLLVPVYKTGKSYSPFGGSNK	1260						
Qy	1261	NVAGMTIYGDGSPKKMENTALSRYSQKNTFDIIHTQGNDLVRKASVRFAPQDPVPASLN	1320						
Db	1261	NVAGMTIYGDGSPKKMENTALSRYSQKNTFDIIHTQGNDLVRKASVRFAPQDPVPASLN	1320						
Qy	1321	MGSAIGDSDSLTVMENGNIPOITSKYSSDNLAITLHNAAFVRYDGSNGVIRNKQISAMKL	1380						
Db	1321	MGSAIGDSDSLTVMENGNIPOITSKYSSDNLAITLHNAAFVRYDGSNGVIRNKQISAMKL	1380						
Qy	1381	TGVDGKSYQGNAPFIANTVXHYGYSDLGGPITVYNTKQNIASVQGHLMNADYTRLLIL	1440						
Db	1381	TGVDGKSYQGNAPFIANTVXHYGYSDLGGPITVYNTKQNIASVQGHLMNADYTRLLIL	1440						
Qy	1441	TPVENNYARLPPEPSPNTILNTVTVGSKTSDFKKCSYAVDGNNSQGFQIFSSYQSS	1500						
Db	1441	TPVENNYARLPPEPSPNTILNTVTVGSKTSDFKKCSYAVDGNNSQGFQIFSSYQSS	1500						
Qy	1501	GWLDIDTGINNTDITKIVMAGSKTHFTASDHIALSPANGSDAMPYTFKPLEIDASSLAF	1560						
Db	1501	GWLDIDTGINNTDITKIVMAGSKTHFTASDHIALSPANGSDAMPYTFKPLEIDASSLAF	1560						
Qy	1561	TNNIAPLDIIVPETKAKDGRVLGKIQTL SVKRVNYPEDILFLRETHSGAQYMLGVYRI	1620						
Db	1561	TNNIAPLDIIVPETKAKDGRVLGKIQTL SVKRVNYPEDILFLRETHSGAQYMLGVYRI	1620						
Qy	1621	RLNTLLASQLVSRANTGIDITLTMETORLPEPPLGEGFFANFVLPKYDPAEHGDERWFKI	1680						
Db	1621	RLNTLLASQLVSRANTGIDITLTMETORLPEPPLGEGFFANFVLPKYDPAEHGDERWFKI	1680						
Qy	1681	HIGNVGNTGRQPYYSGLMDSSTSMTLFVPYAEGYMHGVRGLGVGYQKITVDNTWES	1740						
Db	1681	HIGNVGNTGRQPYYSGLMDSSTSMTLFVPYAEGYMHGVRGLGVGYQKITVDNTWES	1740						
Qy	1741	APFFYDETQKQFVLINADHDHSGMTQOGIVKNIKKYKGLNVSATGYSAPMDNFSASAL	1800						
Db	1741	APFFYDETQKQFVLINADHDHSGMTQOGIVKNIKKYKGLNVSATGYSAPMDNFSASAL	1800						
Qy	1801	YYWELFYVTPMWCFORLLQEQFDEATQWVNVNYPAGIYVNGEIAPIWNCRPLEETTS	1860						
Db	1801	YYWELFYVTPMWCFORLLQEQFDEATQWVNVNYPAGIYVNGEIAPIWNCRPLEETTS	1860						
Qy	1861	WNANPLDAIDPDAVAQNDPMHYKIATFMRLDQILRGDMAYRELTRDALNEAKWYVRT	1920						
Db	1861	WNANPLDAIDPDAVAQNDPMHYKIATFMRLDQILRGDMAYRELTRDALNEAKWYVRT	1920						
Qy	1921	LELLGDPEPDYGSQWAAAPSLSGAASQTVQAAYQODLTMLGRGGSVKNLRTANSVLGLFL	1980						
Db	1921	LELLGDPEPDYGSQWAAAPSLSGAASQTVQAAYQODLTMLGRGGSVKNLRTANSVLGLFL	1980						


```
QY 1981 PEYNALTDYWGOTLRLRNLNHNLSIDGQPLSLAIYAEPDTPKALLTSMVQASQGSVAV 2040
Db 1981 PEYNALTDYWGOTLRLRNLNHNLSIDGQPLSLAIYAEPDTPKALLTSMVQASQGSVAV 2040
QY 2041 LPGTLSLVRFPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTLLLOQGMELATQSIR 2100
Db 2041 LPGTLSLVRFPVMLERTNLVAQLTQFGTSLLSMAEHDDADELTLLLOQGMELATQSIR 2100
QY 2101 IQORTVDEVDADIAVLAERSRQAQRLEKYQQLYDEDINHGEQRAMSLIDAAAGOSLAGQ 2160
Db 2101 IQORTVDEVDADIAVLAERSRQAQRLEKYQQLYDEDINHGEQRAMSLIDAAAGOSLAGQ 2160
QY 2161 VLSIAGVADLVPNVFLGACGSRWGAAALRASASVMSLSATASQISADKISRSEAYRRRR 2220
Db 2161 VLSIAGVADLVPNVFLGACGSRWGAAALRASASVMSLSATASQISADKISRSEAYRRRR 2220
QY 2221 QBEWELQONADGEVKOMDAQLESKIRREAAQMOVETOQAHTQAOLELLQKFTNKA 2280
Db 2221 QBEWELQONADGEVKOMDAQLESKIRREAAQMOVETOQAHTQAOLELLQKFTNKA 2280
QY 2281 LYSWMRGKLSAIYYQFDTQSFCFLMAQEARRELTDNGVTFIRGGAWNGTTAGLMAGET 2340
Db 2281 LYSWMRGKLSAIYYQFDTQSFCFLMAQEARRELTDNGVTFIRGGAWNGTTAGLMAGET 2340
QY 2341 LLLNLAEMKWLDERALEVTRTVSLAQFYQALSSDNFNTEKLTQFLREGKGVGAS 2400
Db 2341 LLLNLAEMKWLDERALEVTRTVSLAQFYQALSSDNFNTEKLTQFLREGKGVGAS 2400
QY 2401 GNELKLSNRQTEASVRLSDKIFSDYPESLGNTRQOLKQVSVTLPALVCPYEDIRAVLNYG 2460
Db 2401 GNELKLSNRQTEASVRLSDKIFSDYPESLGNTRQOLKQVSVTLPALVCPYEDIRAVLNYG 2460
QY 2461 GSIVMPCSAIALSHGVNDSGQFMDLDFNDNRYLPFEGISVNDSGSLTSPDPAIDROKA 2520
Db 2461 GSIVMPCSAIALSHGVNDSGQFMDLDFNDNRYLPFEGISVNDSGSLTSPDPAIDROKA 2520
QY 2521 LLESLSDIILHRYTIRS 2538
Db 2521 LLESLSDIILHRYTIRS 2538
```

RESULT 5

```
US-10-753-901-14
; Sequence 14, Application US/10753901
; Publication No. US20040194164A1
; GENERAL INFORMATION:
; APPLICANT: Bintrim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Larrinua, Ignacio
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Schafer, Barry
; APPLICANT: Bevan, Scott
; APPLICANT: Young, Scott
; APPLICANT: Guo, Lining
; TITLE OF INVENTION: Xenorhabdus TC Proteins and Genes for Pest Control
; FILE REFERENCE: DAS-105X
; CURRENT APPLICATION NUMBER: US/10/753,901
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,717
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 14
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-753-901-14
```

```
Query Match 42.1%; Score 5496.5; DB 4; Length 2523;
Best Local Similarity 44.7%; Pred. No. 0;
Matches 1168; Conservative 436; Mismatches 833; Indels 175; Gaps 36;
QY 7 LLNKISPTRDGGTMTLADQLVLSFELRKKIFDDQLSWGEARHLHYHTIEQKKNRLLEAR 66
```

```
Db 7 LLNKINRKRSGDTLLLTNISFMSFSEFRHRTSGTLTWRETDFLYQQAQESKQNKLEBLR 66
QY 67 IFTRANPOLSGAIRLGIERTDSVSRSDMGFARSSSVKPGSVASMPSPAGVLTLEYREA 126
Db 67 ILSRANPOLANTWNITPSTLNNSYNWFGRAHRFVKPGSIASIFSPAAYLTLEYREA 126
QY 127 KDLHSSAYHLNRRPDLADLTLSQSNMDTEISTLTLSNELLLEHITRKTGGSDALME 186
Db 127 KDFPDNSQYHLNKRPPDIASLALTQNMDEISTLSNELLHNIQTLEKTDYNGVMK 186
QY 187 SLSTYRQADTPYHQPIETIROVIMTHDSTLSALSRNPVWQAGASILLAILANISPEL 246
Db 187 MLSTYRQGMTPYHLPYESARQALLQDKNITAFSRNTDVAELMDPTSLLAIKTIDISPEL 246
QY 247 YNILTETTEKNADALPAQNF--SENITPENFASOSWIAKYKYLELSEVQKYLQMLQNGYS 305
Db 247 YQILVEEITPENSTELMKCNFTDDVL--IFKSVASLARYYDLSYDELSLFWN--LSFCKK 303
QY 306 DSTSAYVDNISTGLVNNESKLEAYKIRVKTDDYDKNINYPFDLMYEGNQQFFIRANPKV 365
Db 304 NTNQYKNEQLITLVNDGNDTATARLIKTRKDPYDSHLNTAELIPIKENEY--KYNFSV 361
QY 366 SREFGATL--RKNAGPSGIVGSLSGPLIANTNPKSNVLSISDSEYKNGVKIYAYRTS 422
Db 362 KKTBPDLDFLQNGDKKEYIYQDNFVPIANTHY--SIPIKLTTEQITNGITLRLWRVKP 419
QY 423 STSATNCGGFTFESVPLTIFALKNAIKRLCTLSGLSPNELOTIVRSDNAQGIINDSV 482
Db 420 NPSDAINNAYFKWMEPGDIFLLKNAIKRLYKATGISPEDIQVIESIYDDLTIDSNV 479
QY 483 LTKVFTYLFYSHRYALSPDDAQVLNGSVINQYADDDSVSHFNRLFNTPPLKGIKEADGN 542
Db 480 LGKLPYQYVQYMOHYNISVS DALVLCHSDISQYSTKQPSHFTILFNTPELLNGQESADNT 539
QY 543 TVSIDPBEQSTFARSALMRGLGVNSGELYQGLAGVLDAQNTITLSVFISSLYRLTL 602
Db 540 KLDLTPGESKNHFYLGIMKRAFRVNDTELYTLWKLANG--GTNPFCMCSIENLSLLYRVL 598
QY 603 LARVHOLTVALNCLMGLSPENGTASTLSGELPRLVIMLYQVTQWLTEAITTEAIWL 662
Db 599 LADIHLITVNLMSLLSVSPVNTKIALFSDTALTQLISFLPQCTQWLITKQWSVDVPL 658
QY 663 LCTPEFGNISPETISNLNNLRPSISEDMAQSHNRLOAEILAPFIAATLHASPDMARY 722
Db 659 MTTDNYSTVLTPDIENLITLNSGLS--TSLGDELIRAA--AAPLIAASIQMSAKTAE 715
QY 723 ILLWTDNLRPGGLDIAGFMVLVKESLNANETTLQVQCHYMAQLSLSVQTLRLSEAEIS 782
Db 716 ILLWINGIKPQGLTFDDFMIIAANRDRSENETSNMVAFCQVLGQLSLIVRNIGISELT 775
QY 783 VLVISGPAVLGAKNOP-----AGQHNIDTLFSLYRFHQWINGLGNPGSDTLDMLRQQ 834
Db 776 LLV-----TKPEKFOSETTALQHDLPQLALTFHAVIMRCGSYATEILLTALSELG 825
QY 835 TLTDADRLASVGLDISWYQAM--VSAGVNOQCQDINTVQLQMDIVASALHTWPSVIRT 892
Db 826 ALTAQLAVALKFDQAVVQALQQTGLGVNTFTNWRITDVTQLQMDVAATLGITPDGVAA 885
QY 893 LVNIRYVTLNKAESNLPWDDEWQTLAENMEAGLSTQQAQTLADYTAERLSSVLCNWFLA 952
Db 886 LILKLYI---GEPETPMPTFDDWQAATLLQAGLNSQQSDQLQAWLDEATTAAASYIK 942
QY 953 NIQPEGVLSHSRDLVSYFLIDNQVSSAIKTRRLAEAIAGQLVINRNLNRIEPPNARADV 1012
Db 943 NSAPQ--QIKSRDELSYLLIDNQVSAQVKTTRVAEAIASQLVYNRNLNVEGKVPKVP 1000
QY 1013 STROFFTDW--TVNNRYSTWGGVSELVYPENYIDPTQRIGQTRMDELENNISQKLSRD 1071
Db 1001 KTRQFFCDWETYNRRYSTWAGVSELAYYPENYIDPTIRIGQTRGMNLLQOLSQSLNID 1060
QY 1072 TVEDAFTYLTFRFETVADLKVVSAYHDVNSNTGLTWFGVGTRENLPELYKRVNDISRMQ 1131
```

Db 1061 TVEDSPKNYLTAFEDVANLQVTSYGHDSINYNVNEGITYLIGYSQTEPRYYWRNVHDQKQC 1120
Qy 1132 AGEIAAANAKWTKIDTAVNPYKDAIRPVIIFRERHLIWEKEBAKNGTDPVETVDRFT 1191
Db 1121 HGQFAANAWGKKTEIPINWQENIRPVIYKSRYLILWLEKELKNESEDKIDITDYI 1180
Qy 1192 LKLAFLRHGDSWSPMSYDITTOVEAVTDKPPDTERLALASGGQGEDTLVAVVYKTKGS 1251
Db 1181 LKLSHIRYDGSWSPFNFNVDKLENLKNKAS---IGMYCSDSEYKDVIIIVFHEKKDN 1237
Qy 1252 YSDPGSKNKNVAGMTIYGDGSKKMENTALSRYSQKNTFDIITQGNLDLVRKASYRPAQ 1311
Db 1238 YS--FNSLPAREGMTINPDWTLISILTENDLD--AIVKSTLSELDTRTE--YKVNQFAT 1290
Qy 1312 DF-----EVPASLNMGSAGIG---DDSLTVWNGNI-----PQTSKYSSN----- 1349
Db 1291 DYLAEBKESITTKNKLASFTGNIPDLSYISPCNGHINLTFNPSMEINFSGKNIYNDEVKY 1350
Qy 1350 -LAITLHNAATFVRDGSNGVIRNKQISAMKLTGVDSKSOYGNAFIANTVKHYGGYSDL 1408
Db 1351 LLSMVEDETIVLFDYDRHEML-----GKEE-----EYFHYG----- 1382
Qy 1409 GGPITVYKNTKNYIASVOGHLMNADYTRRL-----ILTPVENNYIARLFEF 1454
Db 1393 -----TLDFIISID--LKNAEYFRVLMHLRTKEKIPKSEIGVGINVDYESNDAEF 1431
Qy 1455 PFSPTNLTN-----TVFTVGSNKTSPDK-----KCSYAVDGNNS 1488
Db 1432 KLDNTNIVLDKNDTGVWHTICESEFTNDVSIINNMGNTAAALFLREDPCVYLCSATD---- 1487
Qy 1489 QGFOIFSSYQSSGMLDIDTGINNTDITKITVMAGSKTHT--FTASDHASLPANSFPAMPYT 1547
Db 1488 --IKIASSNIEQ-----IODKNISFLKNGSDILVELNAEDHVAKPSHESDPWYD 1537
Qy 1548 FKPLEIDASSLAFTNNTNIAPLDIVFETKAKGRVLG--KIKOTLSVKRVNYPNEDILFRET 1606
Db 1538 FNQVKVDIEGYDI-----PLVSEFTIKQPDGGYNDIVIESPHIKLKSKDTSNLSLHKM 1592
Qy 1607 HSGAQYQMGVYRIRLNTLLASQSVRANTGIDITLWETORLBPPLGEGFFANFVLPK 1666
Db 1593 PSQTQYMQIGFYRTRLNTLFRKLABRANIGIDNVLSTMETQNLBPQLGEGFYATFKLPP 1652
Qy 1667 YDPAHGDERRFKTHIGVNGVGTGPYYSGLMDSITSETMFLVPVYAEYMGHGVRLG 1726
Db 1653 YNKEHGDERFKTHIGINDGNSARQPYEGMLSDI--ETTTLFVPYAKGYIIEGVRLG 1711
Qy 1727 VGYQKTYDNTWESAFFYDETKQOQVFLINDADHDSGMTQQGIKNKIKYKGFNLVSTAT 1786
Db 1712 VGYKKIYDKSWESAFFYDETKNQFIINDADHDSGMTQQGIKNKIKYKGFHVVVMK 1771
Qy 1787 GYSAPMDPNSASALYYWELFYTPMCTORLLOEQFDEATQWYVYNPAGYVNGEIA 1846
Db 1772 NNTPEMDPFGANAIYFWELFYTPMVPQRLLOEQNFTESTRWLYIWNPAYSYQGSQM 1831
Qy 1847 PWIWNCRPLEETSNANPLDAIDPDAVANDPMHYKIATFMRLLDOLILGDMAYRELT 1906
Db 1832 DYNNVNRPLEEDTSNANPLDSVDPDAVANDPMHYKVATFMKLDLITIGDSAYRQLE 1891
Qy 1907 RDALNEAKWYVVRTLELLGDEPDYDYGSOQAAPSLSGAASQTVQAAYQODLTMLGRGVGS 1966
Db 1892 RDTLNEAKWYVQALTLTGDEPYFSLDNDWSEPRLEEAASQTMRRHHYQHKMLQKRAAL 1951
Qy 1967 KNLRTANSVLGLFLPEYNPALTDYQOTLRLRFLNRLNLSIDGQPLSLAIYAEPTDPKAL 2026
Db 1952 PTKRTANSLTALFLPQINKLQGYQWTLTORLYNRLNLTIDGQPLSLSLYATPADPSML 2011
Qy 2027 LTSVMQASGGSAVLPGTILSVYRFPVLMERTNLVAQLTQFCTSLLSMAEHDDADDELFTL 2086
Db 2012 LSAAITASQGGGDLPHAVMPWTRFVPILENAKWGSQILIQFNTLLSITERQDAEALAEI 2071
Qy 2087 LLOQGMELATOSIRIQORTVDRVDADIIVAFSRSQAQRLEKYQOOLYDEDINHGEQAM 2146
Db 2072 LQTCQSELQSIKMQKMAEIDADKLALQSRHGAQSRFDSFNTLYDEDVNAGEKQAM 2131

Qy 2147 SLLDAAAGQSLAGQVLSTIAEGVADLVNPFVFLACGSRGWGAALRASASVMSLSATASOYS 2206
Db 2132 DLYUSSSVLSTSGTALHMAAADLVNIFYGVAGSRFGALFNASAIIGTISASATRIA 2191
Qy 2207 ADKISRSEAYRRRRQEWIQRDNADGEVKQVDAQLESKIRREAAQMVYQETQQAHTQ 2266
Db 2192 ADKISQSEIYRRRRQEWIQRNNAEAEIKQIDAQLATLAVRREAAVLQKNYLETQQAQTQ 2251
Qy 2267 AQLLELQKNTKALYISWMRGKLSAIYYQFEDLTQSCFMAQEAALRELTDTNGVTFIRGG 2326
Db 2252 AQLAFLQSKFTSNAALYVNLGRLSAIYYQFYDLAVSLCLMAEQTYQYELNNAAAHFIRKP 2311
Qy 2327 AWWGTTAGLMAGETLLINLAEMEKVWLERDERALEVTRTVSLAQFYQALSSDNFNLTLEK 2386
Db 2312 AMHGTYAGLAGETMLNLAQMEKSYLEKBERALEVTRTVSLAEVYAGLTENSILKDKV 2371
Qy 2387 TQFLREGKNGVAGSNGELKLSNRQIEASVRLSDUKIFSDYSPESLGNTROLKQVSVTLPAL 2446
Db 2372 TELVNAGEGSGAGTTLNGINVEGTQLQASLSKLSDLNIATDYPDGLGNTERRIKQISVTLPAL 2431
Qy 2447 VGPYEDTRAVLNYGGSIWMPRGCSAIALSHGVNDSGQFMDLPNDSRYLPPREGISVNDSGS 2506
Db 2432 LGPYQDVRALISYGGSTMPRGCKAIAISHGMNDSGQFMDFNDAKYLPPFEGLPVADTGT 2491
Qy 2507 LTLSPFPDATROKALLSLSLDIILHRYTIRS 2538
Db 2492 LTLSPGIGSKQKSLLSLSLDIILHRYTIRS 2523

RESULT 6
US-10-754-115-14
; Sequence 14: Application US/10754115
; Publication No. US20040208907A1
; GENERAL INFORMATION:
; APPLICANT: Hey, Timothy
; APPLICANT: Schleper, Amanda
; APPLICANT: Bevan, Scott
; APPLICANT: Birtim, Scott
; APPLICANT: Mitchell, Jon
; APPLICANT: Li, Ze Sheng
; APPLICANT: Ni, Weiting
; APPLICANT: Zhu, Baolong
; APPLICANT: Merlo, Don
; APPLICANT: Apel-Birkhold, Patricia
; APPLICANT: Meade, Thomas
; TITLE OF INVENTION: Mixing and Matching TC Proteins for Pest Control
; FILE REFERENCE: DAS-104XC1
; CURRENT APPLICATION NUMBER: US/10/754,115
; CURRENT FILING DATE: 2004-01-07
; PRIOR APPLICATION NUMBER: US 60/441,723
; PRIOR FILING DATE: 2003-01-21
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: Patent in version 3.2
; SEQ ID NO 14
; LENGTH: 2523
; TYPE: PRT
; ORGANISM: Xenorhabdus nematophilus
US-10-754-115-14

Query Match 42.1%; Score 5496.5; DB 4; Length 2523;
Best Local Similarity 44.7%; Pred. No. 0;
Matches 1168; Conservative 436; Mismatches 833; Indels 175; Gaps 36;

Qy 7 LNKISPTRDQQTWTLADQLVLSSELKRIKIFDDQLSWGEARHLIYHETIEOKKNELLEAR 66
Db 7 LLDKINKRSQDITLLTNISFMSFSEPHRTSGTLTWRETDFLYQQAQESKQNKLESLR 66
Qy 67 IFRANPOLSCAIRLIGTERDSVRSYDEMFGARSSSVFKPGSVASMPSPAGYLTLYREA 126
Db 67 ILSRANPOLANTNLTTPSLNNYSNWFVGRHFRVFKPGSIASIFSAPYALTYLXREA 126
Qy 127 KDLHFSSSAYHLNRRPDLADLTLSQSNMOTETISTLTLSNELLLEHITRKTGSGSDALME 186

; SEQUENCE CHARACTERISTICS:		
; LENGTH: 2516 amino acids		
; TYPE: amino acids		
; TOPOLOGY: linear		
; MOLECULE TYPE: protein		
US-10-262-794A-47		
Query Match	38.7%; Score 5043; DB 4; Length 2516;	
Best Local Similarity	42.6%; Pred. No. 0;	
Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;		
Qy	22 LADQLYLSFSLRKFIDFQDLSGCRARHLYHETIBQKNNRLLLEARIETFRANPOLSGAIRL 81	
Db	21 LTDLSHSFNFRRQVSEHLSWSETHLDYHDAQQAQKQNRLEYARILKRAPQLQNAVHL 80	
Qy	82 GIERDSVSR-SYDEMFGARSSFFVKPSVASMFSPAGYLTLYREAKDLHPSSAYHLDN 140	
Db	81 AILAPNAELIGYNNQFSGRASQYVAGTVSMFSPAAAYLTLYREARNLHASDVYVYLD 140	
Qy	141 RRPDLADLTLSQNMWTEISLTLSNELLLEHI-TRKGGSDALMSLSLYRQAIDTPY 199	
Db	141 RRPDLKSNALSQNMWTEISLTLSNELLLESIKTESKLENYTKVMEMLSTFRPSGATPY 200	
Qy	200 HOPVETIROVTMTDSTLSALSRNPEVMGQAGASLLAILLANISPELYNILEITEKNA 259	
Db	201 HDAYENVREVLOQDPGLEQLNASPATAGLMHQASLLGINASISPELFNILEITEGNA 260	
Qy	260 DALFAQNPSENITPENFASQSWIAKYGYGLELSEVQKYLGMLOQ-GYSDSTSAYVDNISTG 318	
Db	261 EELYKKNFG-NIEPASLAMPYLYKRYNLSDEELSQPTGKASNFQQEYSN---NQLITP 316	
Qy	319 LVVNESKLEAYKTR-VKTDYDKNINYPDLMEYEGNQFFIRANFKVSRFGAT----- 372	
Db	317 VNNSDGTGVKYRITREYTNAYQMDVELFP--FGEN---YRLDYKFKPNFYNASLYSIK 371	
Qy	373 -----LRKNAGPSGIVSLSGPLIANTNPKSNVLSNISSEYKNGVKIYAVRYTSTS 425	
Db	372 LNDKRELVRTGAQ-----VNI EYSANITLNTAB--ISQPEIGLTVLPSSG 418	
Qy	426 ATNQGGGIFTPESEPLTIFALKANKATRLCUTSLGSPNELTIVRSNQAQIINDSVLTK 485	
Db	419 WA-YAAAKFTVEEYNOYSLKLNKALRSRATLSPTILSGIVRSVNLQDINTDVLGK 477	
Qy	486 VFYTLFYSHRVALSFDDAQLNGSVINQYADDSVSHENLFTNPPLKGIPEADGNTVS 545	
Db	478 VFLTKYMYQRYAIHAETALICNAPISQSRVDNQPQDFLNTPLNGQFPSTGDEID 537	
Qy	546 IDPBEQSTFARSALMRGLGVNSGELYQLGKLAGLDAQNTITLSVFISSLYRLTLIAR 605	
Db	538 LN-SGSTGDWRKTLKRAFNDVSLFRLALKITDHNKDGKIKNLKNLSNLYIGKLLAD 596	
Qy	606 VHLQTVNELCMLYGLSPFNGKTTAS-LSSGELPRLVIMLYQVOTWLTAETITTEAIIWLC 664	
Db	597 IHQLTIDELDLLL-IAVGEKTNLSAISDKQLATLIRKLNITITSLWHTQKMSVQLPIMT 655	
Qy	665 TPFSGNISPEISNLNLRPSI---SEDMAQSHNRELOAAILAPFIAATLHLASPDWAR 721	
Db	656 STSYNKTLPTRIKNLLDTVYHGLQFDKQKAD-----LILHVMAPYIAATLQLSSENAH 709	
Qy	722 YILLWTDNLRPGGLDIAGFMT-----LVLKESLNANETQO-----LVQFCHVMAQLSLS 770	
Db	710 SVLLWADKLQGD-----GANTAEKFWDLNTKYTPGSSEAVEYTOEHIVQYQALQALEMV 765	
Qy	771 VQTLRLSEALSVLVISGFAVLGAKNQAGQHNIDTFLPSLFRHQWINGLGNPGSDTLDM 830	
Db	766 YHSTGINENAPRLFTVKE-EMFGAATGAAPAHADALSILMLTRFADWNVALGEKASSVLAA 824	
Qy	831 LRQQTLTADRLASVMGLDISWYQAMVSA-----GVNQLQCODINTVLQWIDVA 880	
Db	825 FEANSLTAEQADAWNLNANLLOQASIOAQNHQLHPVPTPENAFSCWTSINTILLQWNVVA 884	
Qy	881 SALHTMPSVIRTLVNIIRVYVTAINKAESNLPSWDEWQTLAENNEAGLSLQQOACTLADYTA 940	
Db	885 QQLNVAPOQVSALVGLDYIQSMKET-----PTYAQWENAGAAGVLTAGLNSQQANTLHAFIDE 940	
Qy	941 RLSSVLCNWFANIQPEGVLSHSRDDLYSYFLDNDVOUSSAKTRTLAEATAGIOLYNRA 1000	
Db	941 SRSAALSTYYIRQVAKAAAAIKSRDDLYQYLLDNDQVSAAKTRTRIAEASIASIOLYNRA 1000	
Qy	1001 LNRIEPNARADVSTROPFFTDW-TVNNRYSTGGVSRVLYYPENYIDPTORIGOTRMDEL 1059	
Db	1001 LENVEENANSGVLSRQPFIDWDKYNKEYSTWAGVSQLVYYPENYIDPTMRIGOTKMDAL 1060	
Qy	1060 LENISQSKLSDTVEDAFKTYLRFETVADLKVVSAYHDNWNSTGLTWFGQOTRENLEPE 1119	
Db	1061 LQSVSQSLNADTVEDAFMSYLTSEFQVANLKVISAYHDNINNDQGLTYFGLSETDAGE 1120	
Qy	1120 YYRNVNDISRMQAGELAAWAKWTKIDTAVNPYKDAIRPVIRERLHLTWVEKEEVAK- 1178	
Db	1121 YYRNVSDHSPKDFKFAANAWSEWHKIDCPINPYKSTIRPVYKSRLYLLWLEQKEITKQ 1180	
Qy	1179 --NGTD--PVETDYDRPTLKLAFRLRHGDSWSPMSYDITTOVEAVTDDKPDTERLALAASG 1234	
Db	1181 TGNKSGQYQYETDYRYELKLAHRYDGTWTPITFDVNNKISL--KLEKNRAPGLYCAG 1238	
Qy	1235 FQGEDTLVFPYTKGYSDFGSGKNVAGMTIYVGDGSKKMENTALSRLSOLKNTFDII 1294	
Db	1239 YQGEDTLVNFYNNQDLDYSY--KNASMQGLIYIPADMASKDMTPEQSNVYRD--NSYQQF 1294	
Qy	1295 HTQGNDLAVRKASVYRPAQDFEVPASLWNGS--AIGDDSLTYMENGNIPOITSKYSSDNLAI 1352	
Db	1295 DTNN---VRRVNNRYAEDYEIPSVSRKDYGMGYDYLWVYNGDIPITINKAASSDLKI 1351	
Qy	1353 -----TLNNAFTVRYDGSNVIRNKQISAMKLTGVDSGKSYGNAPFIANTVKHYGY 1405	
Db	1352 YISPLRIIHNG-----YEQ-----KENQCNLMNKG-----KLGDKFIV-----Y 1388	
Qy	1406 SDLG-GPITVYNTKYN-IASVQGHLMNADYTRRLIITPVENNYIARLFEF-PPSPNTIL 1462	
Db	1389 TSLGVNPNNSNKLMPYQVYSGNTSGLNQGR--LLFHRDITFTYPSKEAWIPAKESLT 1446	
Qy	1463 NTVFTVGS-----NKTSPFKCSYAVDGNNSQGFQIPSSYQSSGMLDIDTGINNTDIK 1515	
Db	1447 NQNAAGDDVATDSLNKPDLLKQYIFMTD---SG-----TATDVSGPVEINTALSPAKVQ 1499	
Qy	1516 ITVMAGSKTHTFTASDHASLAPANSFDAMPYTFKPLEIDASSLAFTNNIAPLDIVFETKA 1575	
Db	1500 IIVKAGGKEQFTTADKDVSIQSPSPFDEMYQFNALBIDGSLFINNSASIDVTFAPA 1559	
Qy	1576 KQGRVLG-----KIQTLVSVKRVNTNPEDILFLRETHSGAQYMLGVTRIRLNTLLASQLV 1631	
Db	1560 EDGRKLGYESFSIPVTLKV-----STDNALTLLHNENGAQYMQWQSYRTRLNTLFAQLV 1614	
Qy	1632 SRANTGIDTILTMETQRLPEPPLGEGFFANFVLPKYPDAEHGDERWFKIHIGNVGGNTGR 1691	
Db	1615 ARATGTIDTILSMETQIQEQLGKGYATFVIPPYLSLTHGDERWFKLYIKHVVDNNS- 1673	
Qy	1692 QPYYSGLMSLDTSETSMTLF-----VPYAEGYTHMEGVRLGVGYQKITVDNTWESAFFYD 1746	
Db	1674 HIIVSGQLTDTN-INIITLPLDDVPLNQDVH-----AKVYMTFKSPSDGTWVGPHFVRD 1728	
Qy	1747 ETQOQFVLINDADHDGSGTQGIIVKNIKVKYGFNLVSIATGYSAIPMDPNSASALYYWELF 1806	
Db	1729 D--KGIVTIN-----PKSILTHFESVNLNNS-----SEPMDFSGANSYFWELF 1772	
Qy	1807 YITPMCMCFORLLQSKQFDEATQWYVYVNPAGYLVNGEIAPIWNCRPLEETTSWNNPL 1866	
Db	1773 YITPMLVAQRLHLEQNFDEANRWLKYVWSPSGYIVHGQIQYQNNVRPLEEDTSSNSDPL 1832	
Qy	1867 DAIDPDAVAQNDPMHYKIATFMRLLDQLILRGDMAYARELTRDALNEAQWVYVTRLELGD 1926	
Db	1833 DSVDPDAVAQNDPMHYKVSTFMRLLDQLIARGDHAYRQLERDRLTNEAQWVYVTRLELGD 1892	
Qy	1927 BPEYDGSQQAAPSLSGAASQTVQAAYQODLTMLRG-----GVSKNLRTANSIVGLFPEY 1983	
Db	1893 KPYPPLSTWSDPRDLRAADITTONAHDISAIVLRQNIPTPAPISLRSANLTLTDLFLPQI 1952	

1984 NPALTDYQTLRLRLNRLNLSIDGQPLSLAIYAETDPKALLTSMVQASGGSAVLPG 2043
1953 NEVMANYQTLAQRVYNRLNLSIDGQPLSLAIYAETDPKALLTSMVQASGGSAVLPG 2012
2044 TSLYRPPVMLERTRNLVAQLTQFTSLLSMAEHDADELITLLLOQGMELATQSIROQ 2103
2013 FMSLWRFPHMLERTRNLVAQLTQFTSLLSMAEHDADELITLLLOQGMELATQSIROQ 2072
2104 RTVDSDVADIAVLESRSQAQRLSKYQOQYDDEINHGQRAMSLDAAAGOSLAGQVLS 2163
2073 KTIIELEDAKTVLEKSKAGAGRPDSYGLKLDENINAGENQANTLRASAAGLTTAVQASR 2132
2164 TAEGVADLVPNVFLGACGSRGWAALRASAVMSLSATASQVSAQSDKISRSEAYRRRQEW 2223
2133 LAGAAADLVPNIFGAGGSRGWAALRASAVMSLSATASQVSAQSDKISRSEAYRRRQEW 2192
2224 EIORNADGKQMDAQOESLKRIRERAAQMVQYQETQOATQAOLELLOKFTNKALYS 2283
2193 EIORNADGKQMDAQOESLKRIRERAAQMVQYQETQOATQAOLELLOKFTNKALYS 2252
2284 WNRGKLSAIYQFPLTQSFCLMAQEARLRELTNGVTFIRGGAWNGTTAGLMAGETILL 2343
2253 WLRGELAIIYQFPLTQSFCLMAQEARLRELTNGVTFIRGGAWNGTTAGLMAGETILL 2312
2344 NLAEMKVLDERDALEVRTVSLAQFYQALSSDN--FNTEKTLQFLREKGNVQASG 2401
2313 SLAQMEDAHLKRDKRALEVERTVSLAEVYAGLPKNGPFLSAQEDKLVSGSGSAGSN 2372
2402 NELKL-----SNRQTEASVRLSDIKI FSDYPESLGNTRQLKOVSTLPAALVGPYEDIRAV 2456
2373 NNAFAGAGTDTKTSLSQSVSFADLKIRIDYPSALGKIRIRIKQISVTLTPALLGPYQVQAI 2432
2457 LNYGSIWVPRCSAIALSHGVNDGQFMDPNDKSYLPFEGISVNDGSIITLSPDAT- 2515
2433 LSYGDKAGLANGECEALASHGVNDGQFMDPNDKSYLPFEGISVNDGSIITLSPDAT- 2491
2516 ---DRQKALLESLSIILHRYTIR 2537
2492 PEKGGQATMLKTLNDIILHRYTIR 2516

RESULT 9
US-10-706-424-2
; Sequence 2, Application US/10706424
; Publication No. US20040103455A1
; GENERAL INFORMATION:
; APPLICANT: french-Constant, Richard
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens
; FILE REFERENCE: 62878
; CURRENT APPLICATION NUMBER: US/10/706,424
; CURRENT FILING DATE: 2003-11-12
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-706-424-2

Query Match 38.7%; Score 5043; DB 4; Length 2516;
Best Local Similarity 42.6%; Pred. No. 0;
Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;

22 LADLOYSFSEURKIFDDQLSGWEARHLYHETIEOKNRLLEARIFFRANPOLSGAURL 81
21 LTDISHSFNEPRQCVSEHLSWSETHDYHDAQQAQKONRLYEALIRANPOLQNAVHL 80
82 GIERDSVSR-SYDEMFGARSSSVKPGSVASWMSFSPAGVLTLYREAKDLHFSSSAYHLDN 140
81 AILAPNAELIGNYQFSGRASQYVAPGTVSSMFSPPAALTELYREARNLHASDSVYILD 140

141 RRPDLADLTLSQSNWDTBISTILTSNELLEHI--TRKTGGSDALMESLSTYRQADITPY 199
141 RRPDLKSNALSQONMDIELSTLSNELLESIKTESKLENYKYVMEMLSTFRPGATPY 200
200 HOPVETRIQVIMTHDSTLSALSRNPEVMQAGASLLAILANISPELVNLTETETKNA 259
201 HDAYENREVILQDQPLEQNASPAIAGLHQASLLGINASISPELVNLTETETKNA 260
260 DALFAQNFSENITPENFASQSWIAKYGLEISEVQKYLGMQLN-CYSOSTSAVDNISTG 318
261 EELYKKNFG-NIEPASLAMPVLYKRYNLSDELSQIFGKASNFGQOBSYN---NQLITP 316
319 LVNNESKLEKAYKITR-VKTDDYDKNINYFDMVEGNQOFPFIRANFKVSRREGAT----- 372
317 VVNSDGTGVKYRITREYTNAYQMDVELFP--FGGEN---YRLDYTKNFYNFNSYLSIK 371
373 -----LRKNAGPSGIVGSLGGLIANTNPKSNLSISSEYKNGVKIYAYRYTSSTS 425
372 LNDKRELVRTGAFQ-----VNIEYSANITLTAD--ISQPEIGLTVLPQSG 418
426 ATNOGGGIFTFESYPLTIFALKLNKAIKRLCTGLSPNELQTVIRSDNAQGIINDSVLTK 485
419 WA-YAAAKFTVEEYNOYSFLLKNAIRLSRATESLPTILEGIVRSVNLQDLINTDVLCK 477
486 VFYTLFYSHRYALSPDDAQVLNGVINQYADDSVSHFNRLPNTPLPKGKIFEADGNTVS 545
478 VFLTKYMQRYAIHAETALILCNAPISQISQSDYNDQSFQDFLNTPLNGQYESTGDEED 537
546 IDPBEQSTPARSALMRGLGVNSGELYQLGKLAGVLDAQNTITLSVFISSLYRITLLAR 605
538 LN-SGSTGDWRKTYLKRAFNTDDVSLFRLLKITHDNKDGKIKNNLKNLSNLYIGKLLAD 596
606 VHQLTVNELCMLYGLSPFNGKTTAS-LSSGBELPRVILWYQVQWLTBEAITEATEILLC 664
597 IHQLTIDELDLA-LAVGEGKTNLSAISDKQLATLIRKLTNTITSMHTQKNSVOLFIMT 655
665 TPEFSGNISPRISNLLNLRPSI---SEDMAQSHNRLOQAEILAFIATLHLASPDMAK 721
656 STSYNKTLPBPKNLLDVTYHGLQGFQDKAD-----LLHVMAPYIAATLQLSSENVAH 709
722 YLLATWDLNLRPGGLDIAGFMT-----LVLESINANETQO---LVQFCHVMAQLSLS 770
710 SVLLWADKLQFGD-----GAMTAKEFWDLNNTKTPGSSSEAVETQEHIVQYCOALAQLKV 765
771 VQTLRLSEALSVLVIGFAVLGAKNOPAGQHNIDTFLSYRHFHOMINGLNGPQSDTLD 830
766 YHSTGINENAPFLFVTKP-EMFGAATGAAPADALSLIMLTFADWVNAVGEKASVLA 824
831 LRQOTLTADRLASVWGLDISWQAMVSA-----GVNQLQWQDINTVLQWIDVA 880
825 FEANSLTAEQLADANLNDANLLQASIQAHQHQLPPVTPPENAFSCWTSINTILQWVNA 884
881 SALHTMPSVIRTVANIRVVTALNKAESNLPSWDEWQTLAENWEAGLSQOAOITLADYAE 940
885 QQLNVAPOGSALVGLDYIQSMKET-----PIYAQWENAGVLTAGNLSQOANTLHAFDE 940
941 RLSSVLCNFWLANTQPEGVSLHSDLYSYFLIDNQVSSAIKTTTLRAEAIAGIQLYINRA 1000
941 SRSAALSTYIRQVAKAAAAIKSRDLYQLLIDNQVSAAIKTTTRIAEAIASIQLYVNA 1000
1001 LNRPEPNARADVSTQFPTDW-TVNNRYSTWGVSRVLYYYPENYIDPTQIGOTRMDEL 1059
1001 LENVEANANGVIRSQFFIDWDKYNKRYSTWAGVSQLVYYPENYIDPTMTRIGOTKMAL 1060
1060 LENTSQSKLSDTVEDAFKTYLTRPETVADILKVSAYHDNVNSNTGLTFWVGOTRENLEPE 1119
1061 LQSVSQSNQADTVEDAFMSVLTSEQVANKVISAHDNINNOQGLTYFGLSETDAGE 1120
1120 YYWRNVDIRMQAGEAANAWKEWTKIDTAVNPKDAIRPVIFFERLHLIWEKEEVAK- 1178
1121 YYWRSVHDSKFNCKGFAANANSEWHKIDCPINPKYSTIRPVIKSRLLYLLWLEQKEITKQ 1180
1179 --NGTD--PVETDRFTLKLAFRLRHGDSWAPMSYDITTOVEAVTDKPKDTERLALASG 1234

Db 1181 TGNKSGVQVETDVRVELKLAHRYDGTWNTPIFDVNNKISL--KLEKNRPFGLYACAG 1238
Qy 1235 FQGBDILLVYVYTKGKSYDFGGGNKNVAGWTIYGDGSKFQKQENTALSRYSQLKNTDII 1294
Db 1239 YQGBDILLVYVYTKGKSYDFGGGNKNVAGWTIYGDGSKFQKQENTALSRYSQLKNTDII 1294
Qy 1295 HTQGNDLVRKASYRPAQDFEVPASLNWGS--AIGDSDITVMENGINIQTISKYSDNLAI 1352
Db 1295 DTNN---VRRVNNRYAEYIEPSSVSRKQYGVGDYLLSMVYNGDIPNTYNAKASSDLKI 1351
Qy 1353 -----TLHNAAFVRYDGSNGVIRNKQISAMKLTGVDGKSOYQGNAPFIANTVKHYGGY 1405
Db 1352 YISPKLRIHNG-----YEQG---KRNQCNLMNKG-----KLGRKFIV-----Y 1398
Qy 1406 SDLG-GPIVYVYTKKTY--IASVQGHLMNADYRRLILTFPVNNYIARLFEF-PPSPNTIL 1462
Db 1389 TSLGVNPNNSNKLMPYVYQYSGNTSLNQR--LLFHRDTTPSKVEAMIPGAKERSLT 1446
Qy 1463 NTVTIVGS-----NKTSDPKKCSYAVDGNNSQGFQIFSSYQSGWLDIDTGNNTDIK 1515
Db 1447 QNAAIGDDYATDSLNPDDLUQYIFMTD-----SKG---TATDVGSPVEINTALISPAKVQ 1499
Qy 1516 ITVMAGSKTHFTTASDHIASLPANSFDMPTFKFPLEIDASSLAFTNNAIPLDIVFETKA 1575
Db 1500 IIVKAGGKEQFTADKQVSIQSPSFEDMNYQFNALEIDGSLFINNSASIDVTFATA 1559
Qy 1576 KQGRVLG-----KIQTLISVKNVNNPEDIPLFRETSGAQVQMGVYRIRLNTLLASQV 1631
Db 1560 EDGRKLGYSPISFVTLKV-----STDNALTLHNEGAQVQWQVSYRTRTLTLFARQLV 1614
Qy 1632 SRANTGIDTILMTETORLPEPLGSGFPANVLKYDPAEHGDERWFKIHGNVGGNTR 1691
Db 1615 ARATGIDTILMETQIQEPQLGKFVATVIPPYNLSTHGBRWFKLYIKHVVDNNS- 1673
Qy 1692 QPYTSGMLSDTSETSMTLF-----VPYAEGYMHGVRGLGVGYQKITVDNTWESAFFVD 1746
Db 1674 HIIYSGQLTDN-INIITLFIPLDDVPLNQDH---AKVYMTFKPSDGTWGWPHFVRD 1728
Qy 1747 ETQOQFVLINDADHDSGWTQGIKVKIKYKGFNLVSIATGYSPMDPNSASALYYWELF 1806
Db 1729 D--KXIVTIN-----PKSILTHFESVNVNLNIS-----SEPMDFSGANSLYFWELF 1772
Qy 1807 YITPMWCFORLLOKQFDEATQWYVYVNPAGYLVNGEIAPIWNCPCLETTTSWNAPL 1866
Db 1773 YITPMLVAQRLHQBQFDEANRWLYVWSPSGYIVHGQIQYQNVNRPLEDTSWNSDPL 1832
Qy 1867 DAIDPDVAQNDPMHYKIATPMLLDQILILGDMAYRELTLDALNEAKWYVYRLELIGD 1926
Db 1833 DSVDPDAVAQNDPMHYKVSTPMTLIDLLIARGDHAYQLERDITLNEAKWYVYRLELIGD 1892
Qy 1927 EPEDYGSQWNAAPSLGASQSTQAAVQODLTMLGRG---GVSKNLRTANSLVGLFPEY 1983
Db 1893 KPYPFLSTWSDPRLDRAADITQNAHDSAIVALRQNIPTAPILSLASANTLTDLFLPQI 1952
Qy 1984 NPALTDYQWTLRLRLFNLRHNSIDGQPLSLAIYAEPDPKALATSMVQASQSGSAYLPG 2043
Db 1953 NEVMNMYQTLAQVYNLRHNSIDGQPLSLAIYAEPDPKALATSMVQASQSGSAYLPG 2012
Qy 2044 TLSLYRPFVMLERTNLVAQLTQFGTSLLSMAEHDDADDELITLLIQQWELATQISIRIQ 2103
Db 2013 PMSLWRPFHMLNARGVWSQLTQFGSTLQNIIEHQDAENALLQNAQAEILLTSLSD 2072
Qy 2104 RTVDEVDADIAVLAEBSRRSAQRLEKYQQLYDVEDINHGEBQAMSLDAAAGQSILAGQVLS 2163
Db 2073 KTIELDAEKTVLEKSKAGASQSRFDSYKLYDENINAGENOAMTLRASAAGLTAVQAQR 2132
Qy 2164 IAEGVADIVPNVGLACGSGRWGALRASAVMSLSATASQYADKISRAEYRRRQEW 2223
Db 2133 LAGAADLVPNIFGAFAGGSGRWGALRASAVMSLSATASQYADKISRAEYRRRQEW 2192
Qy 2224 EIQRNADGEVKQMDAQLESILKIRREAAQMVQETOQQAHTQAEQLLQKFTNKALYS 2283

Db 2193 EIQRNNAEALKQIDAQKSLAVRREAAVLQKTSLKQEQOTSQLAFQKFSNQALYN 2252
Qy 2284 WMRSKLSAIYYQFPDLITQSPCLMAQEAALRRELITDNGVTIFRGGAWNGTTAGLMAGETLL 2343
Db 2253 WLGRGLAAIYFPQFYDLAVARCLMAEQAYRWELNDDSFARFKPGAWQGTYAGLAGAETJML 2312
Qy 2344 NLAEMEKVWLERDERALERTVTRTSLAFOYQALSSDN--FNLTEKLTQFLREGKGNVQASG 2401
Db 2313 SLAQMEDAHLKRDKALEVERTVSLAEVYAGLPKXNGPFSLAQEDIKLVSGSGSAGSGN 2372
Qy 2402 NELKL-----SNRQIEASVRLSDLKIFSDYPESIGNTRQLKQVSVTLPALVGPVEDIRAV 2456
Db 2373 NNLAFAGTDTKTSLOASVSFADLKIREDPASLGKIRRIKQISVTLPALILGPYQDVQAI 2432
Qy 2457 LNYGGSIVMPRGCSAIALSHGVNDSGQFMDLFDNDSRYLPEGISVNDSGSITLSFPDAT- 2515
Db 2433 LSYGDKAGLANGEALAVSHGMDSGQFQDFNDGKLPFEGIAI--DQGTLTLSFPNASM 2491
Qy 2516 ---DROKALLESLSDIILHRYTIR 2537
Db 2492 PEKGQATMLKTLNDIILHRYTIK 2516

RESULT 10
US-10-703-280-4
; Sequence 4, Application US/10703280
; Publication No. US2005012713A1
; GENERAL INFORMATION:
; APPLICANT: Dow Agrosciences LLC
; TITLE OF INVENTION: USE OF UNTRANSLATED REGION OF OSMOTIN GENE TO ENHANCE TRANSGENE
; TITLE OF INVENTION: EXPRESSION IN PLANTS
; FILE REFERENCE: 61,795A
; CURRENT APPLICATION NUMBER: US/10/703,280
; CURRENT FILING DATE: 2003-11-07
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 4
; LENGTH: 2516
; TYPE: PRM
; ORGANISM: Photorabius luminescens
US-10-703-280-4

Query Match 38.7%; Score 5043; DB 5; Length 2516;
Best Local Similarity 42.6%; Pred. No. 0;
Matches 1110; Conservative 460; Mismatches 837; Indels 198; Gaps 57;

Qy 22 LADQYLSFSELKRIFFDQDLSWGEARHLHYHETIQKXNRLLEARIPTANPQLSGAIRL 81
Db 21 LTDISHSFNEFRQOVSEHLSWSETHLDYHDAQQAQKXNRLYEALILKXNRLPQLQNAVHL 80
Qy 82 GIERDSYR-SYDEMFCARSSSFVKGSVASMSPAGYLTREYKADHESAYHLN 140
Db 81 AILAPNAELIYNNQFSGRASQYVAPGTVSMFSPAAYLTREYKARNLHASDVSYYLDT 140
Qy 141 RRPDLADTLISQSNMDTEISTLTLSNELLLEHI--TRKTGGSDALMESLSTYRQAITPY 199
Db 141 RRPDLKMSALSQQNMDIELSTLSNELLLESIKTESKLENTYKVMELSTFRSGATPY 200
Qy 200 HQPYETIRQVIMTHDSTLSALSARNPEVMQAGASLAILANISPELWILTEITEKNA 259
Db 201 HDAYENREVIQLQDQGLEQLNAPAGLHQAQALGINASISPELFNLTITEBTEGNA 260
Qy 260 DALFAQNFSENIPTENFASQSWIAKYKYLELSEVQKYLGMQLQ--GYSDSISAYVDNISTG 318
Db 261 EELYKKNFG-NIEPASLAMPYELKRYNLSDEELUSQFGKASNFGQGEYSN---NQLITP 316
Qy 319 LVVANSKELAYKITR-VKTDYDDKINIFYDLMYEGNQQFFIRANFKVSRFEGAT----- 372
Db 317 VVNSDGTVKVYRITREYTNAYQMDVELFP--FGGEN---YRLDYKFKFNYSYLSIK 371
Qy 373 -----LRKNAGPSGIVGSLGGLIANTWFKSNYLSNISDSEYKNGVKIYVAYRTSTS 425
Db 372 LNDKRELVRTEGAPO-----VNIEYSANITLNTAD--ISQPFELGLTRVLPSSG 418


```

QY 1667 YDPAEHGDERWPKIHGNVGNTRGPYYSGMLSDTSETMTLFPVY---AEGYYMHGV 1723
Db 1644 YDESHGTNKSFAIEVDIFKENDSFVYIYQGLSETSTVVKVFLSYFTEATGNKNLWV 1703
QY 1724 RLGVGYKITYNTWESAFYFDETKQO-----FVLINDADHDSGMTQOQIVKNIKKYKGF 1779
Db 1704 R--AKYQKETTDK-----ILFDRDDEKDPHGFWSLDDHKTFSGLSSAQAALN----- 1748
QY 1780 LNVSIATGAPMDFNSALYYWELFYTFMFCQRLLOEQFDEATQWQWNVVNPAGY 1839
Db 1749 -----DSEPMDFSGANALFYWELFYTFMMAHRLLOEQFDAANHFVYVWSPGY 1800
QY 1840 IVNGSIAPWNCRLPLETTSWNAFLDAIDPAVQNDPMHYKIATFMRLLDQLILRGD 1899
Db 1801 IVDGKIATYHWNVRPLEEDTSWNAQQLDSDTPDAVAQDDPMHYKVATFMTATLIDLWARGD 1860
QY 1900 MAYRELTDALNEAKWVYRTLELGDPEDEYGSQWAAAPSLGSAASQTVAAYOODLTM 1959
Db 1861 AAYRQLERDYLAEAKWVYTOALNLLGDBFQVMLSTTWTANPTLGNAAKTKTQOVROOVLQ 1920
QY 1960 LCRGVSKN--LRTANSVLGFLPEYNPALTDYWTOLRLNLRNLSIDQOPLSLAIY 2017
Db 1921 LRLNSRVKTPPLGTANSALTALFPOENSKLGYWETLAQRMENLRNLSIDQOPLSLPLY 1980
QY 2018 AEPTDPKALLTSMVQSGSVAFLPGTSLYRFPVWMLBTRNLVAQLTOFTGTSLLSMAEH 2077
Db 1981 AKPADPKALLSAVASASQGGADLPKAPLTTHRFPQMLEGARGVNLQIFGSSLLGYSER 2040
QY 2078 DDADELITLLQOQNELATQSIQRIQRTVDEVDADIIVLAESRRSAQNRLKYQOYDED 2137
Db 2041 QDAEAMSQLTOQASSELITLSIRMDNQLAELDESEKLTALQVSLAGVQORFDSYSQLYEEN 2100
QY 2138 INHGQRAMSLDDAAAGOSLAGOVLSIAEGVADLVNVPVFLGACGSRWGAALRASVMS 2197
Db 2101 INAGQRALALRESAIESQNOISRMAGAGVDWAPNIFGLADGMHYGATAYATADGIE 2160
QY 2198 LSATASQYSADKISSEAYRRRRQWEIQRONADGEVKQMDAQLBSLKIRREAAQMVVEY 2257
Db 2161 LSASAKWVDAEKAQVSEIYRRRRQWEIQRONADGEVKQMDAQLBSLKIRREAAQMVVEY 2220
QY 2258 QETQAAHTQAAQLLELQRTKALYSWMRGLSAIYQFFDLTQSFCLMAQEAALRELTLD 2317
Db 2221 LKTOQAQAQAQTFLRSKFSNQALYSWLRGLSGIYQFYDLAVSRCLMAEQSYQWEAND 2280
QY 2318 NGVTPRGAWNGTTAGLMAGETLLNLAEMEKVWLDERDALEVRTVSLAQFYQAL-S 2376
Db 2281 NSISFVKFGAQGTAGLLCGEALLQNLQAQEEAYLKWESRALEVERTVSLAVVYDSLEG 2340
QY 2377 SDNFMNLEKLTQFLREGKNGVSGASGNELKLNROIEASVRLSDLKIFSDYPESL---GNT 2433
Db 2341 NDRFNLAEOIIPALLDKGEGTAGTKENGKLSANAILASVGLSDKLGTDYDPSIVGSNKV 2400
QY 2434 ROLKQVSTLPAIVCPYEDIRAVLYNGSIVNPRCSAIALSHGVNDGSGQFMLDNDERY 2493
Db 2401 RRIKQISVSLPALVQYQDVQAMLSYGGSTQLPKGCSALAVSHGNTDNGSQFQDLPNDGKY 2460
QY 2494 LPFEGISVNDGSLTLPFDATDROKALLESLSLILHRYTIR 2537
Db 2461 LPFEGIALDDQSTLNLQFPNATDKQAILQTNMSDILHRYTIR 2504

```

RESULT 12

```

US-10-262-794A-12
; Sequence 12, Application US/10262794A
; Publication No. US20030207806A1
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: French-Constant, Richard

```

```

Query Match 38.2%; Score 4986.5; DB 4; Length 2504;
Best Local Similarity 42.3%; Pred. No. 0;
Matches 1110; Conservative 434; Mismatches 863; Indels 217; Gaps 48;

```

```

QY 9 NKISPTRDG-----QMTLADQLYLSFSELAKIFDDQLSWGEARHLVHTTEQKKNRL 63
Db 3 NSLSSTIDTICQKLQLTCPABIALYPPFTFREKTRGMVNWGEAKRIYIAQAEQDRNLLH 62
QY 64 EARIFTPRANPOLSGAIRLGIERD-SVRSRSDYDEMFGARSSSVKPGSVASMFSPAGYLT 122
Db 63 EKRIFAYANPLLNKNAVRLGTQMLGFTIQGYSDLPFNADNYAAPGVSASFSPAAIYLT 122

```

```

; APPLICANT: Rocheleau, Thomas A.
; APPLICANT: Blackburn, Michael B.
; APPLICANT: Hey, Timothy D.
; APPLICANT: Merlo, Donald J.
; APPLICANT: Orr, Gregory L.
; APPLICANT: Roberts, Jean L.
; APPLICANT: Strickland, James A.
; APPLICANT: Guo, Lining
; APPLICANT: Cliche, Todd A.
; APPLICANT: Sukhapinda, Kitisri
; TITLE OF INVENTION: Insecticidal Protein Toxins From Photorhabdus
; NUMBER OF SEQUENCES: 88
; CORRESPONDENCE ADDRESS:
; ADDRESS: Dow Agrosciences Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/262,794A
; FILING DATE: 02-OCT-2002
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/851,567
; FILING DATE: 05-MAY-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27386
; REFERENCE/DOCKET NUMBER: 960296.93804
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2504 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-10-262-794A-12

```


Qy 1286 QLKNTEDIHTQND-----LVKASVFAODEEVPASL--NMGSAGDSDSLTVMXGNI 1338
 Db 1271 NSYQFDFVMADPDSNDKVKVTRRVNNRYAEIDYELPSSVTSNSYSGDHSLSLWLYGGSV 1330
 Qy 1339 PQITSKYSDSLAITLHNAAFVTVRVDGSGNIRNKQISAMKLTGVDGKSGYGNFIANT 1398
 Db 1331 PNITFESAEDRLST-NWALSIIHNGYAGT-RRIQCNLMK----- 1369
 Qy 1399 VKHYGGYSDLGPIITVYNTKRYIASVOGHLNADYTRRLILTVPVNNYARLPEFPSP 1458
 Db 1370 -----QYASLGDKFTIYDS-----SFDANRPNLVP-----LPFGKDE 1403
 Qy 1459 NT-----ILNT-----VFTVGSNKTDFPKCSYAVD-GNNSQGFQ----- 1492
 Db 1404 NSDDSICTYNNPSEDKKWFSSKDDNKTDYNGGTCQCIDAGTSGKDFYNNLOEIEVIS 1463
 Qy 1493 ----IFSSYQSGWLDIDTGINNTDITKTVMAGSKTHTFTA--SDHIASLPANSFDMAPY 1546
 Db 1464 VTGGWSSYKISNPININTGDSAKVKYTVKAGGDDQIFTDADNSTYVPQPPAPSEMIY 1523
 Qy 1547 TFKPLEIDASSLAFTNNIAPIDIVFETKAKGRVLGKIQTLVKRVNYPEDILFRET 1606
 Db 1524 QFNNTIDCKNLNFIDNQAHIEIDPTATAQDGRFLGAETFIIPVTKVLGTENVIALYSE 1583
 Qy 1607 HSGAQYMLGVYRIRLNTLLASQLSVSRANTGIDTILMETORLPEPPLGEGFEANFVLPK 1666
 Db 1584 NNGVQYMOIGAYRKLNTLFAQLVSRANRGIDAVLSMETQNIQEPQUGAGTYVQLVLDK 1643
 Qy 1667 YDPAEHGDERWFKIHIGNVGNGTGRQPYYSGLMDSMTSETSLFVY---AEGYMHGV 1723
 Db 1644 YDESHTGNTKSPAIEYVDI KRENDSFVIYQGESETSQTVVKVFLSYFIETATGNKHLWV 1703
 Qy 1724 RLGVQYKITYDNTWESAFFYDETKQ-----FVLINDADHDGWTQOQIVGNI KYKGP 1779
 Db 1704 R--AKYQKETDK-----ILFDRDDEKPHGFWLSDHDKTFESGLSSAQLN----- 1748
 Qy 1780 LNVSTATGYSAPMDPNSALYYELFVYTPMCFQRLLOEQKQDEATQWNNVYNPAGY 1839
 Db 1749 -----DSEPMDFSGANALYFELFYTPMMAHRLLOEQNFDAANHFYVWSPGY 1800
 Qy 1840 IVNGEIAPIWNCRPLETTSSWNAFLDAIDPDVAQNDPMHYKIATPMRLDQLILRGD 1899
 Db 1801 IVDGKIALYHNVNPLEEDTSWNAQLDSTDPDAVQDDPMHYKVATPMATLIDLMLARGD 1860
 Qy 1900 MAYRELTDLALNEAKWVYRTELLELGDPEPDYGSQOWAAPSLSGAASQTQAAVQOOLTM 1959
 Db 1861 AAYRQLERDTLAEAKWYTOALNLLIGDEPQVMLSTTANPTLGNAAASKTTQOVRQOVLTO 1920
 Qy 1960 LGRGVSKN--LRTANSIVGLFLEPNPALTDYQTLRLRLNLRHNLSDIGQPLSLAIY 2017
 Db 1921 LRLNSRVKTPLLGTANSITLFLPOENSKLKGWYRTLAQRNPNLNRHNLSDIGQPLSLPLY 1980
 Qy 2018 AEPTDPKALLTSMVQASQGSVAFLPGTSLSYRFPVMLERTNLNAQLTQFGTSLLSMAH 2077
 Db 1981 AKPADPKALLSNAVSASQGGADLPKAPLTHRFPQMLGARGLVNQLLQFGSSLIGYSER 2040
 Qy 2078 DDADELTYLLQQGHELATQSIROQRTVDEVDADIAVLESRSQAQRLEKYQOYLDED 2137
 Db 2041 QDAEASQLLOQASELITTSIRMQDNLQALDESEKTLQVSLAGVQOQRFDSYSOLYEN 2100
 Qy 2138 INHGEORAMSLDLAAGOSLAGOVLSIAEGVADIVPNVFLGACGSRGWAALRASAVMS 2197
 Db 2101 INAGEORALARSESAIESQGAQISRMAGAVDMAPNIFGLADGGMHYGATAYADGIE 2160
 Qy 2198 LSATASOYSAOKISREAYRRRRQEWBIQRDNADGEVQMDAQLESKIRREAAQOMVEY 2257
 Db 2161 LSASAKVDAEKVAQSEIYRRRRQEWKIQRDNAAQAEINQLNAQLESLSIRREAAEMOKEY 2220
 Qy 2258 QETQOATQAOLELLORFETHKALYSWRGKLSAIYQFPDLTQSFCMLMAQEARLRELT 2317
 Db 2221 LKTOQAQAQALTFURSKPSNALYSWLRGLSGIYFQFDYDLAVSRCLMAEQSYQWEAND 2280
 Qy 2318 NGVTFIRGAMNGTTAGLMAGETLLNLNAEMEKNWLERDERALBVRTVSLAQFYQAL-S 2376

Db 2281 NSISFVPGAQOQTYAGLGCCEALIQNLAQMEAYLKWESRALEVERTVSLAVVYDSLEG 2340
 Qy 2377 SDNPNLFEKUTQFUREBKGNVAGSNELKLSNRQIEASVRLSDLKIFSDYPESL---GNT 2433
 Db 2341 NDRNLAEQIPALLDKGEGTAGTKENGSLANATLSASVKLSDLKAGTGDPDSIVGSNKV 2400
 Qy 2434 RQLKQSVTVPALVGPVEDIRAVLNYGSGIVMPRGCSAIALSHGVNDSGOQMLDFNDSRY 2493
 Db 2401 RRIKQISVSPALVGPYQDVQAMLSYGGSTQLPKGCSSALAVSHGTNDSGQQLDFNDGKY 2460
 Qy 2494 LPFEGISVNDSGSLTSLFFPDATDRQKALLESLSLIIILHRYTIR 2537
 Db 2461 LPFEGIALDDQGTNLNLQFPNATDKQKAILQTMDSIILHRYTIR 2504

RESULT 14
 US-11-020-848-11
 ; Sequence 11, Application US/11020848
 ; Publication No. US20050155104A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Apel-Birkhold, Patricia
 ; APPLICANT: Hey, Timothy
 ; APPLICANT: Sheets, Joel
 ; APPLICANT: Meade, Tom
 ; APPLICANT: Li, Ze Sheng
 ; APPLICANT: Lira, Justin
 ; APPLICANT: Russell, Sean
 ; APPLICANT: Thompson, Robin
 ; APPLICANT: Mitchell, Jon
 ; APPLICANT: Fencil, Kristin
 ; TITLE OF INVENTION: Toxin Complex Proteins and Genes from Xenorhabdus bovienii
 ; FILE REFERENCE: DAS-114XCI
 ; CURRENT APPLICATION NUMBER: US/11/020,848
 ; CURRENT FILING DATE: 2004-12-23
 ; PRIOR APPLICATION NUMBER: US 60/534,893
 ; PRIOR FILING DATE: 2004-01-07
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: Patent in version 3.2
 ; SEQ ID NO 11
 ; LENGTH: 2534
 ; TYPE: PRT
 ; ORGANISM: Xenorhabdus bovienii
 US-11-020-848-11

Query Match 38.2%; Score 4976; DB 6; Length 2534;
 Best Local Similarity 41.4%; Pred. No. 0;
 Matches 1089; Conservative 490; Mismatches 861; Indels 190; Gaps 49;

Qy 1 MYSTAVLLNKISPTRDGQMTLADQLVLSFSELRKIPDQLSWGCEARHLYHETIEQKNN 60
 Db 1 MINTENILIRUNRESSQEPMTLAHIMFISFAFRKEVKDTLWNGESHLYLAAKAEKEN 60
 Qy 61 RLLEARIPTTRANPOLSGAIRLGIERSYSRSYDEMFGARSSSVFKPGSVASMFSPAGYLT 120
 Db 61 RIFEARULLSRANPOLRGAVRLIGIQQLSORQSYDTLFGGRSGKYLPGSVASMFSPAAYLT 120
 Qy 121 ELYEAKDLHPSSAYHLNDRPDLADLTLSQSNMDTEISTLTLSNELLEHI-TRKGG 179
 Db 121 ELYESRHLHSSESIYHLDKERPDLSQIMLTQENQDQTLSTLSLNSDLDFGINKKKLN 180
 Qy 180 DSDALMESLSYRQADITFPYHOPVETTRQVIMTHDSTLSLSRNPENVMQAGEASLLAIL 239
 Db 181 KNEVDLKMWSLDRWLSGNTFPYHQPETLSNIVSQDLPQLSQVSPKVLGLLSVPSLLGIS 240
 Qy 240 ANISPELYNLTETITEKNADALFAQNFSENIPTENFASQSWIAKYGLEISEVQKYLGM 299
 Db 241 SQISPELYKILTEBITAENADQMYKQFGD-LPISALSNPNYLMKYDIDADTLRAVMGI 299
 Qy 300 LQNGYSSTSAVVDNISGLVNNESKLEAYKITRVKTDYDDKNIYFDLAVYEGNNQFFI 359
 Db 300 YGSGQNDDEPAFISDQAIVTYLLDDKNGSFVTYLTIRTKGETYDMQVNFTEIPTKDGKUKY 359

Db 2424 AARIKQVSVSLPALLGPVQVQAILSYSGDMKIGPKGCSALAVNGMNDSSQFQOLDPND 2483

QY 2491 SRYLPFGISV---NDSSGLTSLFPDADTRDQKALLESIDILHRYTIR 2537

Db 2484 TKYLPFGINIPKDKQDQALVSPFNADAKQKTMLLSLDILHRYTIR 2533

RESULT 15

US-10-706-424-8

; Sequence 8, Application US/10706424

; Publication No. US20040103455A1

; GENERAL INFORMATION:

; APPLICANT: fFrench-Constant, Richard

; APPLICANT: Waterfield, Nicholas

; TITLE OF INVENTION: DNA Sequences from tcd Genomic Region of Photorhabdus luminescens

; FILE REFERENCE: 62878

; CURRENT APPLICATION NUMBER: US/10706,424

; CURRENT FILING DATE: 2003-11-12

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 8

; LENGTH: 2381

; TYPE: PRT

; ORGANISM: Photorhabdus luminescens

US-10-706-424-8

Query Match 36.8%; Score 4802; DB 4; Length 2381;

Best Local Similarity 40.8%; Pred. No. 0;

Matches 1053; Conservative 449; Mismatches 793; Indels 284; Gaps 48;

QY 22 LADQLYSFSELKIFDDOLSGEARHLYHETIEOKNRLLEARIETFRANPOLSGAIRL 81

Db 21 LTDICHYSFNRQVSDHLSEWSETRNRYRDRQEQENQLYEARILKRANPOLQNAVHL 80

QY 82 GIERDSVS-RSYDEMFGARSSSFVKPGSVASMFSPAGYLITELYREAKDLHFSSAYHLDN 140

Db 81 GITLPHAEELRGVNSFEGGRASQYVAPGVSVMFSPAAYLITELYREARNLHASDSVYHLDE 140

QY 141 RRPDLADLTLSQNMDSITSLTNELLLEHI-TRKTCGSDA-LMESLSTYRQAITDP 198

Db 141 RRPDLQSMTLQNMDSITSLTNELLLEHI-TRKTCGSDA-LMESLSTYRQAITDP 200

QY 199 YHOPVETITROVIMTHDSLSALSRNPEYVQAGASLLAILANISPELYNLTITEIKN 258

Db 201 YHDAVENVRKAIQLOPKLEQFQKSPAVAGLHQASLLGINNISPELYNLTITEIKN 260

QY 259 ADALPAQNFSENITPENPASQWIAKYGLSELSEVQKYLGLMQLONGYSDSTSAVDNISTG 318

Db 261 AEAIVKQNFQD-IDPACLAMPEYLKSYNFSDEELSQFI-----RKYPDN----- 304

QY 319 LVVNNEKLEAYKITRVKTDDYDKNINFDLMYEGNNQFFIRANPKVSREFGATILRNAG 378

Db 305 -----ELNTO----- 309

QY 379 PSGIVGSLGGLIANTNPKSNVLSNISDSYKNGVKIYAVRYSTSTSATNQGGIFTES 438

Db 310 ----- 309

QY 439 YPLTIFALKKNAIRLCITSGLSPELQTIIVRSNDAQIINDSVLTIFYFLFYSHRYAL 498

Db 310 ---KHLKINKIKILLQAVNLPFLKLEIPEQN---ITPTVLGKFLVLYYQKYN 362

QY 499 SFDDAQLVNGSVINQYADDSVSHENRLENTPLPKGKIFEADGNVTISDPEEOSTFARS 558

Db 363 GTETALILCNDSISQSYSNQSPQDRLENTSPLNGQYFVIEDTNDISLNSDND-WHKA 421

QY 559 ALMRGLVNGSGLYQGLAGVLDQAQNTITLSVFVSSLYRLTLRLARVHQLTVNELCML- 617

Db 422 VLKFAFNVDLISLRLLHIANHNTDGIANNIKLSNLYTKLADLHQITIDELYLILL 481

QY 618 ---YGLSPFNKGTASLSSGELPRVLVIMLYQVQWMLTEAEITTEAIWLCTPEFSGNISPE 675

Db 482 ITIGEDKIN---LYDIDDKELEKILNRLOTLSNMLHTQKWSIYQLFLMTTNYDKLTPE 538

QY 676 ISNLLNLRPSISDMQAQSHNRELOAEILADPFIATLHLASPDMARYILLWTDNLRPGGL 735

Db 539 IONLLDTVYNGL-QNFQKNTKGLAA--IADYIATIQLPSENVAHSLLWADKPKPEN 595

QY 736 DIAGFMILTILVLESNANETT-----QLVQFCHVMAQSLSVQTLRLSEALSVLVTS 788

Db 596 KITAEKFWLQNRDITELSKPEPMEQEQIIOYCHLAQLTWIYRSSGINENAFRIEKP 655

QY 789 --FAVLGAKNQPAQOHNIDITLSLYRPHOWINGLNGPQSDTLDMLRQOTLTADRLASVNG 846

Db 656 TIFGIPDEPNKATPAHNAPTLIILTRFANWVNSGKASPLITAFENKTLTAEKLANAMN 715

QY 847 LDISMVTOAMVSA-----GVNQLQWODINTVLQWIDVASALHTMPSVIRTLVNIYV 899

Db 716 LDANLEQASIQAYQYKQVTKENTFSNWSQSDIILQWNTIASNLNISQGISPLIALDYI 775

QY 900 TALNKAESNLPSEWQOTLAENMEAGLSTQAOQTLADYTABERLSSVLNCWFLANIQPEGV 959

Db 776 ---KPAQKTPTYAQWENAAIALTAGLDTQOQTHLHVFLDESSTALSNNYIGKVANRAA 831

QY 960 SLHSRDDLYSYFLDNQVSSAIKTRRLAEATAGIOLYINRALNRIEPNARADVSTROFFT 1019

Db 832 SIKSRDDLYQVLLIDNQVSAEIKTRIAEATIASQLYVNRALENIETHAVSDVITROFFI 891

QY 1020 DW-TVNNRYSTWGGVSVLVPYIPYIDPTORIGOTRMWDELLENISQSKLSDTVEDAFK 1078

Db 892 DWDKYNKRYSTWAGVSQLVYIPYIPYIDPTMIGOTKMMDTLQSVQSQQLNADVEDAFK 951

QY 1079 TYLTRFETVADLVKVSAYHDVNSNTGLTVPVGOTRENLPEYVNRNDVISMQAGELAA 1138

Db 952 SYLTSFEQVANLEVISAYHDVNSNTGLTVPVGOTRENLPEYVNRNDVISMQAGELAA 1011

QY 1139 AWKQWTKIDTAVNPKYKAIKRVPIFRERHLIWEKEEVAQNGTDPVETD-RFTLKLAFL 1197

Db 1012 AWSEWHKIDCAINPYQSTIRPVYKSLRYLTLWLBQKETAQKEDNKVTTDYHELKLAH 1071

QY 1198 RHDCWSAPWSYDITTOVEAVTDDKPTDLAALASGFGEDTLLVYVYKTKGYSDFGG 1257

Db 1072 RYDGTWNVPIFDVDEKILALELYK--SQAPGLCAGYQGGEDTLLIIMFYRKKEKLDDY-- 1127

QY 1258 SNKXVAGMTIYGDGSKFKKMENTALSRYSQLXNTFDIHTQGNDLVRKASYRPAQDFEVA 1317

Db 1128 KTAQMOCFYIFSDMSSKDMTNEQCNRYDNGYTHFTDNTSDNSVIR-INNRVADYEIPS 1186

QY 1318 SLNMGSA--IGDSSLTVWENGNIPOITSKYSSDNLAITLHNAAFTRYDGSNVRNKQI 1375

Db 1187 LINSNSHDWGEYNLSQVYGGNI-VINYKVTNSDLKIYI----- 1224

QY 1376 SAMKLTGV-DGKSQVGNAFIANTVGHYGG-----YSDLG-----GPITVY 1415

Db 1225 -SPKRLIHDQKE--GREIOSNLIKKGKLGKDFIYTSGLINPNSSNPFMFVYVQY 1281

QY 1416 NKTNYIASVQGHLMNADYTRRLILTPVENNYIARLFEF-PFSPNTILNTVTVGS---- 1470

Db 1282 NGNTSGLA--QGRLL---FHR-----DTSYSKVAAWIPGAGKSLINENANIGDDCAE 1329

QY 1471 ---NKTSDFKKCSYAVDGNNSQGFQIFSSYQSSGWLIDITDITNTDITKITWAGSKHTFF 1527

Db 1330 DSVNKPDDLQYIYMTD---SKG---TATDVSGPVDINTAISSEKVOITIKAG-KEYSL 1381

QY 1528 TASDHASLPANSPDAMPYTPKPLEIDASSIAFTNNIAPLDIVFETKADGRVLG-KIKQ 1586

Db 1382 TANKDVSQVSPSFEEMCYQFNALIEDGSNLFNTNNSADIVTFTALADDERKGLGYELFN 1441

QY 1587 TLSVKRYNPNEDILFLRETHSGAQYQMGVYRIRLANTLLASQLVSRANTGIDTILTMET 1646

Db 1442 IPVQKV--KTDNALTLFHDENGAQYQWQYRIRLNTLFAQLQVERANTGIDTILSMET 1499

QY 1647 QRLPEPPLGEGFFANFVLPKYDPAEHGDERWFKTHIGNVGNTGQRPQYSGMLSDTSITS 1706

Db 1500 QNIQEPMMGIGAYIELIDKYNPDHGTNKSFKIYIGDIFKAGDHFPIYQOGLSDITQTT 1559

Qy	1707	MTULFVYPAEYGYHHE-GVRLGCVGYOKITYDNTWESAFFYFDETKQOPVLINDADHDSGMT	1766
Db	1560	VKLFPRVDNAYGNKNLXYTAAQYKV-----ETNFIPIVDEKNNKPAT	1603
Qy	1766	QOQIVKNIKKYKFLNVSIAITGVSAPMDFNFSASALYYWELFYITPMFCFORLQAEKOFDE	1825
Db	1604	FDITYKN-GTFPGLASARVLOTUSEPMDFSGNSLYFWELFYITPMVAQELHHEQNFE	1662
Qy	1826	ATOWINYVNPAGYIYVNGEIPAWINCRPIEBETTSSWNAPLDAIDPDAVAQNDPMHYKIA	1885
Db	1663	ANRWLKYVWSPSGYIVRGQIKNTHWNVRPLETNSWNSDPLDSVDPAVAQHDPHYKVA	1722
Qy	1886	TFWRLLDOLITLRGDMAVRELTRDALNEAKMMWYRTELLELGDDEPYDQSOQWAAPSLSGAA	1945
Db	1723	TFWRTLULLMARGDHAHYQROLEDTLNEAKWYMQAHLHGLNGKPYLPLSSVWMDPRJDNAA	1782
Qy	1946	SQTVQAAQQDLMWLRGGVSKN--LRTANSLVGLFLPEYNPALTYDWTQRLRLFLNLRH	2003
Db	1783	ATTQRAHAYITSLRQGTOTPALLARSANTLDTLFLPQINDVYMSYWNKGLRLYLNLRH	1842
Qy	2004	NLSIDGQPLSLAYAEPTDKALLTSMVOQASQGSVLPOTLSLYRPPVMLEKTRNLVAQ	2063
Db	1843	NLSIDGQPLHPIYATPADPKALLSAAVATSQGGKLPESFISLWRPFPMLNARSVMVTO	1902
Qy	2064	LTOFGYSLLSWAHDDELTTLTLQOGMELATQSIRIQORTVDEVDADAVLAESSRSA	2123
Db	1903	LIQFGSTLQNIIEERQDABESLALLQONAKELIUTLUSIQDKTIEEDAEKTVLEKSKAGA	1962
Qy	2124	QNELEKYQQLYDEDINHGEORAMSLMDAAAGQSILAGQV--LSTIAEGVADIVPNVFGLAGC	2181
Db	1963	KSRFDNYSKLYDEDVNAGERQALDM--RIASQSITSLGKLHMAAAALEMVPNIYGFPAVG	2020
Qy	2182	GSRWGAALRASASVMSLSATASQVSADKISRSEAYRRRRQEWIIRDNADGCKVQMDAQL	2241
Db	2021	GTRYGAITANAIAITGGGTAABGLLITAEKVQSQSIWRRRRQEWIIRORNAEAKNQIDAQL	2080
Qy	2242	ESLKIREEAQMOWEYEQTOQAHTQAOQLELLQKFTNKALYSWMRGKLSAIYYQFFDLTQ	2301
Db	2081	KSLTVRREAVALQTKLTKTQOQAOQALFQKRFNSQALYWNLRGLAIIYQFFDLVV	2140
Qy	2302	SFCLMAQEARLREITDNGVTFIRGGAWNGTTAGLMAGETLALLNAEMEKVWMLBERDERALE	2361
Db	2141	ARCLMAEQAYRWETNDSSARFIKPGAWQGTVAGLLAGETLMLNLAQMEDAHLKEQORALE	2200
Qy	2362	VTRTVSLAQYQALSSDNFNLTEKLTQFLREGK-GNVGASGNEKLKLSNRQIEASVRLSDL	2420
Db	2201	VERTVSLAQYVQSIGESKFPALDKIEALLQGDKETSAQNDGNQULKTNNNTLSATLTLQDL	2260
Qy	2421	KIFS DYPE--SLGNTRQOLKQVSTPLALVGPYEDIRAVLNYGGSIV-MPRGCSAIALSHG	2477
Db	2261	KLKDDPIEENMQLGKTRIKQISVSLPALLGYPQDVQAVLSYGGDATGLAKGCKALAVSHG	2320
Qy	2478	VNDSGQFMLDFNDSRYLPPFGISVNDSSGSIUTLGPFDATDQKALLSLSLSDIILHIRYTI	2536
Db	2321	LNDSGQFOLDNFQDGKFIPIFGSIDINDKGTTLTGFPPNAASQKQNILOMTDIIILHIRYTI	2379

Search completed: February 16, 2006, 21:54:06
Job time: 244.217 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:46:54 ; Search time 16.3347 Seconds
(without alignments)
2208.600 Million cell updates/sec

Title: US-10-754-115-34
Perfect score: 13043
Sequence: 1 MYSTAVILNKISPRDQGM.....KALLESDIILHRYTIRS 2538

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 107819 seqs, 14214640 residues

Total number of hits satisfying chosen parameters: 107819

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications_AA_New*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pdb*
- 2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pdb*
- 3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pdb*
- 4: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pdb*
- 5: /cgn2_6/ptodata/1/pubpaa/US05_NEW_PUB.pdb*
- 6: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pdb*
- 7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pdb*
- 8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pdb*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5043	38.7	2516	6	US-10-647-956A-2
2	4986.5	38.2	2504	6	US-10-647-956A-8
3	246	1.9	5024	6	US-10-793-626-2964
4	222	1.7	2710	7	US-11-051-453-41
5	222	1.7	5291	7	US-11-052-554A-281
6	221.5	1.7	2902	7	US-11-052-554A-91
7	202.5	1.6	3194	7	US-11-052-554A-90
8	201.5	1.5	2367	7	US-11-051-453-42
9	192	1.5	2399	6	US-10-995-561-776
10	191.5	1.5	2399	7	US-11-052-554A-92
11	182	1.4	2053	7	US-11-013-759-9
12	178.5	1.4	1889	7	US-11-102-476-46
13	170.5	1.3	2087	7	US-11-075-185-28
14	167.5	1.3	2769	7	US-11-113-424-14
15	165	1.3	2340	7	US-11-052-554A-171
16	160	1.2	1461	7	US-11-052-554A-283
17	160	1.2	1960	7	US-11-069-834-50
18	158.5	1.2	1242	7	US-11-013-711-110
19	157	1.2	1382	7	US-11-098-686-10934
20	155	1.2	980	7	US-11-052-554A-17
21	155	1.2	2668	7	US-11-124-368A-215
22	153.5	1.2	1296	6	US-10-615-668-3
23	153.5	1.2	1449	7	US-11-052-554A-237
24	152.5	1.2	2015	7	US-11-052-554A-374
25	152	1.2	5406	6	US-10-995-561-774

ALIGNMENTS

RESULT 1

US-10-647-956A-2
; Sequence 2, Application US/10647956A
; Publication No. US20050251878A1
; GENERAL INFORMATION:
; APPLICANT: French-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/10/647,956A
; PRIOR FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514
; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
; US-10-647-956A-2

Query Match	38.7%	Score 5043;	DB 6;	Length 2516;
Best Local Similarity	42.6%	Pred. No. 0;		
Matches 1110;	Conservative 460;	Mismatches 837;	Indels 198;	Gaps 57;
Qy	22	LADQYLSFSELKIFDDQLSGEARHLYHTTIQKNNRLLEARIFFRANPQLSGARL	81	
Db	21	LTDISHSSENFROQVSEHLSWSETHDLYHDAQKQDNRLYEARILKRNAPQLQNAVHL	80	
Qy	82	GIERDSVSR-SYDEMFGARSSSVKPGSVASMFSPAGVLTLYLREAKDLHFSSSAYHLDN	140	
Db	81	AILAPNAELIYNNQFSGRAGVAPGVSNFPAAYLTLYLREARNLHADSIVYLDLT	140	
Qy	141	RRPDLADITLSQSNMDTEISTLTLSNELLLEHI-TRKTGGS DALMESLSTYRQADITPY	199	
Db	141	RRPDLKMSALSQNMDELSTLSNELLLESIKTESKLENTYKVMELSTFRPSGATPY	200	
Qy	200	HQPIETTRQVMTHTDSTLSALSRNPEVQAGBSALLAILANISPELNIITEETEKNA	259	
Db	201	HDAYENREVITLQDQPGLEQLNASPAIAGLMHQSLLGINASISPELNIITEETEGNA	260	
Qy	260	DALFAQNFSENIITPENFASQSWIAKYGLSELSEVQKYLGMQLQN-CYSDSTSAVDNISTG	318	
Db	261	EELYKNFG-NIEPASLAMPYLKRYNLSDELSQFICKASNFGQEQEYSN---NQLITP	316	

Db	2373	NNLAFGAGT	TKTSLQ	SAVSFAD	LKREDY	PASLGK	IRRIKQ	ISVTL	PALLG	PGYQ	VOAI	2432
Qy	2457	LNYGGS	VTMPRG	GSAT	ALSHG	VND	SGOF	MLD	FND	SR	LP	2515
Db	2433	LSYGDK	AGLANG	CEAL	AVSHG	VND	SGOF	QD	FND	GK	LP	2491
Qy	2516	---	DRQ	KALL	ES	SD	ILH	RY	TY	IR		2537
Db	2492	PEKG	QATML	KLN	DI	ILH	RY	TY	IK			2516

RESULT 2

```

US-10-647-956A-8
; Sequence 8, Application US/10647956A
; Publication No. US20050251878A1
; GENERAL INFORMATION:
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rocheleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/10/647,956A
; CURRENT FILING DATE: 2003-08-26
; PRIOR APPLICATION NUMBER: CURRENT APPLICATION NUMBER: US/09/817,514
; PRIOR FILING DATE: CURRENT FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 2504
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-10-647-956A-8

```

Query Match	38.2%	Score 4986.5	DB 6	Length 2504
Best Local Similarity	42.3%	Pred. No. 0		
Matches 1110	Conservative 434	Mismatches 863	Indels 217	Gaps 48
QY	9	NKISPTDGG-----QTMTLADLQVLSFSELRKIFDDOLSWGEARHLYHETIEQKKNRL	123	
DB	3	NSLSSTIDTTCQKLQLTCPAEIALYPDPTFREKTRGMVNWGEAKRIEYIAQASODRNLH	62	
QY	64	EARIFFRANPOLGSAIRLGIHERD-SVRSRSYDEMFGARSSSFVKPQGSVASMFSPAGYLTEL	122	
DB	63	EKKIIFAYANPLLNKAVRLGTRQMLGPTQGVSDILFGRNADNYAAPGGSVASMFSPAAYLTEL	122	
QY	123	YREAKOLHFSSAYHLNRRPDLADLTLQSNDMDEISTLTLSNELLLEHITRTKGDS	182	
DB	123	YREAKNLHDSSIIYLDKRRPDLASLMUSQKNWDEISTLALSNELCLAGIEKTKGSD	182	
QY	183	ALMESLSTYRQADTPPHQPYETIRQVIMTHDSTLSALSRNPVWGQAGASILLAILANI	242	
DB	183	EVMDMLSYTLSGETPYHHAYETVREIVHERDEGFRHLSQAPIVAAKLDPTVLLIGSSHI	242	
QY	243	SPELYNLLTTEITEKN---ADALPAONFSENIIPENPASOSWIAKYKLGLELSFVKYLG	299	
DB	243	SPELYNLLIIBEIPEKBAADLTLYKTNFGD-ITTAQMLSPSYLARYYGVSPEDIAYVTT	301	
QY	300	LQN-GYSDSTSAVYDNISTGLVWNNESKLEAYKITRVKTDYDKNINYPOLMTEGNNQFF	358	
DB	302	LSHWGVYS-----SDILVPLVDGVCKMEVVRVTRTPSDNYTSQTNWIELYQGGDNYL	354	
QY	359	IRANPKVRSRFGAT---LRKNAGPSGIVGSLGP---LIANTFNKSNLYLSNISDSEYKNG	412	
DB	355	IKYN--LSNSFGLDLDPYQYKGSADWTEIAHNPPDPMVINKQYESQATIKRSDSD--NI	410	
QY	413	VKIYAVRYTSSATNCGGIGFTFESVPLTIFALKLNKAIARLCUTSGLSNPELQTVRS	472	
DB	411	LSLGLORWHG--GSYNFAANFKIDQSPKAFLLKNNKAIARLKLKATGSLSPATLIERIVDS	468	

Db 1297 NSSNLSSGLYQMAKSVLPDMS-----NLSVSVGTSSIKAINLSQNASINA 1344
Qy 171 -EHITRKTGGD---SDALMESLSTYQAIIDTPYQPIETIROVIMTHDSTLS---ALGRN 223
Db 1345 SNHSTLELQGLDNLNDT---SSLNLQSAINVSNNATINDYASLIASNGSHLNFCAVNFN 1402
Qy 224 PEVMQASGASLAILANISPELYNLTETETKRNADLFAQ-NFSEN-----IT 272
Db 1403 S-----ANITSL---SSSIVFKGAVSLRGOFNLNNSSLDFQSSAIT 1444
Qy 273 P-----ENPASOSWIAKYCYGLELSEVQKYLGMONGYSOSTSAYVDNISTGLVNNES 325
Db 1445 SNTAFNFYDNFASQSPITFHQALDIKVLPLSGGNLNL---PNSSVNLNKNLSQLVFSQDQ 1501
Qy 326 KLEAVKIIRVKTDDYDKNINPDLMEGNNOPIRANFKVSREFCATLRKQAG----- 378
Db 1502 SLNIANIDLLSLNGKRVYNIIOADMNGWYERINP-----FG-WRINDGIYDAKQ 1554
Qy 379 PSGIVGSLGPIJANTFNKSNYLS-----NISSEYKNGVKIYAYRYTSTSTANQGG 431
Db 1555 TVSFNPLNNAIKITESFKNQSLVTLSQLPGIKNTLYNISGEIENYQ-----KVYNNAN 1609
Qy 432 GIFTESPLTIFALKINKAIRLCITSLGSPNELQTIIVRSNQAQIINDSVLTKVPTLP 491
Db 1610 GYVSY-----SDDAQGVYLTSSVKGY--- 1632
Qy 492 YSHRYALGFDDAQVLNGSVINQYADDSDVSHPNRLFTPLKGIKFEADGNTVSDPDEE 551
Db 1633 -----NPNQSYQASGNTTKNNLTSESSVTSQTYNAQNPIS----- 1671
Qy 552 OSTFARSALMRLGVLNGSELVOLGKLAGVLDAQ-NTITLSVFPVTSLSYRLTLARVHOLT 610
Db 1672 -----ALHVVNGYFNFSNIKALQWALKLYPEIKKILGNDFSLSSNLKQDA-LNQLT 1724
Qy 611 -----VNELCMLYCLSPFNCKTTASLSSGELPRIVLWLYQVTLWEABITTEAI 660
Db 1725 KLITPSDMKNINEL-----IDNANNSVQNFNG---TLIIGATKIGQTDNSAVVFG 1776
Qy 661 W-----LLCTPEFSGNISPEISNLNLNRPISSEDMAGSHNRELQAEILAPFIAT 711
Db 1777 GYQKCDYTDIVC-QKPRG-----TYLGQLLESISADLGVIDTFNAKEI---YLTGT 1825
Qy 712 LHLASP-----DMARYILWTDNLRLPGGLDIAGFTWLVLKESLNANETTLQVQF 760
Db 1826 LSGSNAGWTGGSASVTFNSQTSILNQAIVSSQTD--GIFSMQLQEGI--NKVFNQAGL 1881
Qy 761 CHVMAQL-----SLSVQTLRLSEAELSVLVSGEAVLGAKNOP-----AGOHNI 804
Db 1882 ANILGEVAMQSINKAGGLGNLTIVNTLG-----SDSVIGGYLTPEQKNQTLSQLLGQNF 1935
Qy 805 DTLPSLYRPHOWINGLGNPGSDTL--DMLRQO-----TLTADRLASVMGLDISMVTQAMVS 858
Db 1936 DNL-----MNSGLNTAKDLIRQKLGFTWGLVGLAGLGDIDQNPKEKLS 1983
Qy 859 AGVNOQCQDIT-NTVLQWIDVASALHTWPSVIRTLNIRYVTAALNKAESNLPSPDEWOT 917
Db 1984 MSINDLLSKGLFNQITGFISANDIGQVISMQLDIVK-----PS-----D 2024
Qy 918 LAENMEAGLSTQO-AQTLADYTAERLSVLWCWFLANTQPEGVSLHSRDDLYSYFLIDNQ 976
Db 2025 ALKNDVAALGQKMGIEFLGQDTFNSLESLLQ-----OQ 2058
Qy 977 VSSAKTTLRLAIAIGIOLYINRALNRLNTPN-----ARADVSTRQFFTD 1020
Db 2059 IKSVDKVLAKGLSI---YEQGLDLIPNLKKGIFAPYGLSQVQKGFPSFNAQGVN 2115
Qy 1021 WTVNNRYSTWGG-----VSLRVYYPENYIDPTQIGTRMDELLNISQSKLSRDTV 1073
Db 2116 FVQNSTFSNANGGTLFSFAGNSLIFAGNHHIAFTNHSGLTNLLSNQVSNINVTMLNASN- 2174
Qy 1074 EDAPKTYLTRBTVDLKVVSAYHDVNSNTGLTWV----- 1110
Db 2175 -----GLK-INATNNNSVSQ-NLFINASCVCQSDPTTASATNPCTTAQ 2217

Qy 1111 ---GQTRENLPEYYWRNVDIRMQ-----AGELAAANAKWEKTKIDTAVNP-----YK 1154
Db 2218 NNASSSNASNAPIALNNDESIVTANGFNFSNGIYANGVVDFSKI KGSANVKNLLYN 2277
Qy 1155 DAIRPVIREFRLHLIWEKEVAKNGTDPVETYDRFTLKLAFLRHDSWSAPWSDIITQ 1214
Db 2278 NA-----QOANNLTITSNQAVLEKNA-----SEVTNLIIOGAFNNATQK 2318
Qy 1215 VEAVTKKPDTERIALAASFGQGEDTLVFFV--KTGKSYSDFGGSKNV----- 1262
Db 2319 IEVL-----QNLVIA-----SLSTGIYGLEVGALNGLGAIHFNLENSQTPVNP 2366
Qy 1263 -----AGMTIYGDGSKFKOMTALSRY-----SOLKNTFDIHL 1295
Db 2367 IQVGGIINLNTQTPFMNVSVANGGTYYTLKS---SRVIDYNINPNSLQSYLK-LYTLIN 2422
Qy 1296 TQGNDLVRKASYRAQDFEVPASLNMGSAIGDSDLTVMENGINPOITSKYSSDNLAIHL 1355
Db 2423 INGNHIEKNG-----VLTYLQORVLLQDKGLL--SVALPNSNNAQNNILSLSVL 2472
Qy 1356 NAAFTVRVYDGGNVRNKQISAMK--LTGVGDKSOY-----GNAFIANTVVKHYGY 1405
Db 2473 HNQIKMSY---GNKWDFTPTLQDIYVIGIOQSALAQIEAVGGNNAIKWLSTLMWETKE 2529
Qy 1406 SDLGGPITVYKNTKNIYASVQGHLMADYTRLLILTPVENNYARLFPPFPSPNTILNTV 1465
Db 2530 NPLPAPIVLEHNSLNEILGVTKDLQN---TASLISNPNFRNNATSLEW-----ASYTQ 2580
Qy 1466 FTVGSKNTSDPK-----KCSYAVDGNNSQGFIFSS-----YOSSG 1501
Db 2581 QTSRLTKLSDFRAREGESNFSERLELKNKFSDBNPSEVFVKYSQLSKHPNNLWQGV 2640
Qy 1502 WLDDTGTNNNT-----DIKITVMAGSKTHFTTA--SDHIASLSPANSFDMPTFKP 1550
Db 2641 GASPIGSGNGTLYGLNVGYDRLVKSVILGYVAVGYSGFNGNIMHSLANNVDVGYARAF 2700
Qy 1551 LEIDASSLAFTNNIAPLDIVFETKAKOGRVIGKIKQTLV--KRVNYP-----E 1598
Db 2701 LKRNFTLSAN-----ETYGNAHINSSNLSLVLNQRVNYNTWTTTSVNGNYG 2750
Qy 1599 DILFIRET-----HSGAQYMQLV--YRIRLNTLLASQLVSRANTGIDTILTMETORLPE 1651
Db 2751 DFMFKQKSVLKPQVGLSYHFGLSGMKGMQNPAQOQFVHMSPNSBSVLTLMGLBSR 2810
Qy 1652 PPLGEGFPANFVLPKYDPAEHGDERWFKIHGN-----VGGNTGROPYYSGLSDTSETSM 1707
Db 2811 KYFGKNSY-----YFVTARLGRDLLIKAKGDNVVRFGENT--LLYRKGEIFNT----- 2857
Qy 1708 TLFVPAEYGYMH-----EGVRLGVGYQKI 1732
Db 2858 --FASVITGEMHLWRLMYVNAVGLKMLQVQDL 2890

RESULT 7

US-11-052-554A-90
; Sequence 90, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3.
; SEQ ID NO 90
; LENGTH: 3194

Query Match	Score	DB 7	Length	DB 7	Length
Best Local Similarity	18.3%	Pred. No. 0.00029			
Matches	514	Conservative	351	Mismatches	1052
135	AYHLDNRRLDADLTLTQSNMDEITSTLTLNELLLEHITRKTGDSDALMESLSTYRQA	194			
402	AYDDVNNLNIIRLKNQATKONS-KNATSNNTHTYVYVNLGG-----	445			
195	IDTPYHOPVETTRQVIMTHDSTLSALSRNPEVMOQAGASLAIL--ANISPELYNLTTE	252			
446	--TLVH-----FROIPEPSPDILVQSV-----YYGANNLYTNSVNIHDNVEFL--K	487			
253	EITEKNADALPAQN--FSENITPEPASPASWIAKYGLSEVQKYLGMQLONGYSDSTA	310			
488	NIMDRADTFYVNLGLNTWNYTOARPAQT-----YGGKNSA	523			
311	YVDNIST---GLVANNESKLEAYKITRVTCTDDYDKNINYPDLMEYGNNOFFIRANFKVS	366			
524	LVFNAITPWANGAI PKSNS-----TVRPGYB-GVNW-----GKTGYITGTAD	567			
367	REF--GATLRKNAGPSG-----IVG---SLSGPLIAN--TNPKSNYLS-----	402			
568	RVIYITGNMNGCAQTGGGATLNFVGATEINAGATFKNLKTTSONSYMTFALGNGSGS	627			
403	---NISDEY-----KNGV-----KIYARYTSTSA-----TNQGGIPTF	436			
628	GKINVSQSDFYDWTGGYDFTGNGVFDVS VNFNKAY-YKFOGAESYNPKNTFLAGNPKF	686			
437	ESYPLTIFALKNKALRLCLTSLGSPNELOTVIRSDNAQG---INDSVLTKVYTLFYS	493			
687	QG-KTTIEKSVLND-----SYAFDGVNNAFNEKDFNGSGFNPENRAQTNFAWNKSPS	738			
494	HRVALSFDDAQV-----LNGSVINQYADDDSVSHFNLEFNPPLKGKIFEADGNTVSI	547			
739	GG-SFSPNAKQVDFNGSNFNGGVFN-FNNTPKASFTNDENV-----NNQFKIN	785			
548	PDEQSTFARSAL--MRGL-----GV-----NSGELYQL-----G	575			
786	GAQDTFTFGKGVVFNQGLLSLVSCTTYQLLNKASVGVYKDNNAALYQMLRWTSGENPSG	845			
576	KL-----AGVLDQAQNTITLSVFLVS-----SLXRLTLARVHQLTVNELCWL	618			
846	KLVDENKTPASAKIYVNOFTONGLTYYIKENFNNGITLTLCTLGTYH-----CVNI	898			
619	GLSPFNKGKTTASLSSGELPRLVIWLYQVOMLTEAEITTEATLWLTCTPEFSGNISPEIS	678			
899	DNDAFNLKNVNNAS-----NTVFLNGMTWKT-----AGTVFTQDYSSTNSVLVFN	947			
679	-----LNNLRPSISEDMAQSHNRELOAEI-LAPPTAATLHASPDMARYILLTNDLRPG	733			
948	QTPFPFLAGANP--TSNSVVGFGKTSGBEGLVGYIQGVFKANQIDI-----TGTIRSG	998			
734	GLDIAQFWTLVKESINANEITQLVQFCHVAQSLSVQTLRLSEAELSVLVLSG-PAV	791			
999	NGAKTGGGATLV-----FNAQERLNI-----ANANLANDKAGLQNSWNFVNGN	1047			
792	LGA--KNO-PAGQRNIDTLFSLYRPHQW-----INGLGNPGSD-----TLDMLRQ	833			
1048	TNANFSNQTPHGGFNLA-----NNITWDKGSVSGGFGVDNANANGNAVINKVNFSDN	1102			
834	QTLTADRLASVMGLDLSMTVQAMVSAGVN-----QLOCWQDI	870			
1103	GTLLIYKGENSAGNSLTLENNTFNSYNINAKQNLIFNNNSFNSGYSFNDTKNVTFGT	1162			
871	NTVLQWIDVASALHTWPSVIRTLV-----NIRYVTALNKA-ESNLP	910			
1163	NTLIN-SDPFSRLKGSVDNNSIFNIEDRLTKTYYTLGSGDNIKY-----NNQALANVF	1218			
911	SWDEWOTLAEAMEAG--LSTOOAQTLADYAE-----RUSSVLGNWFLANTIOPE	957			

Db 2146 LS--NIVY-----EBSLNDNAIPKDLANNIPKDLGSKTSLSSLLS 2182
 Qy 1871 P-----DAVQN-----DPMHYKIATFMFLDQLIURG 1898
 Db 2183 PTEVNNLLGVSAFKAIMEILNSKTVGVFGENGLLNALDPVKR-----EIDOMLLEQ 2236
 Qy 1899 DMAYRELFDALEAKMYVRLTELLGDEPDYGSQQAAPSLSGAASQTVQAAVQODLT 1958
 Db 2237 IQAHS-----SGEKEPIVKTGLI--ENVENFINNWKQSLSSFANNFVFGGLNQALD 2287
 Qy 1959 MLGRGVSKNLRTANSLVGLPEYNPALTDYQWTLRLRLNRLNLSIDQPL--SLAI 2016
 Db 2288 KIGSSDAKLOQS-----FLDK-----TFGDIL-----NOMINAPLNUKUIS 2326
 Qy 2017 YAEPTDKALATSMVQASQGSVALPGLTSLYRFPVMLERTNLVAQLTQFGTSLSSMAE 2076
 Db 2327 WLGPQDLSVNVIALNSITNPSKELLGAIS-----GNGQKVND-- 2365
 Qy 2077 HDDADELFTLLQOQMEIATOSIRIQRTVDEVDADI-----AVLAESRRSAON 2125
 Db 2366 -----LLGEGVANKIMSNQVLGQWINKIADKGGVYHQGLGSILP--KSLQD 2412
 Qy 2126 PLEKYQQLYDEIDNHGEORAMSLIDAAQOSL--AGQVLSIAGVADLVNVPVGLACGS 2183
 Db 2413 ELKK-----LGMG-----SLKPKGLHNLWQGNFVAKNHVFNNSLFSNATGGE 2459
 Qy 2184 RWGAALRA-----SASVMSLSATASQYSADKISRSSEAYRRRQEWIEIQDN 2229
 Db 2460 LNFVAGKSIIFNGKNTINFTQYQGLSFSVKDFSNISLDTNATNGLTLNASKNDISVQK 2519
 Qy 2230 ADGEYKQMDAQLSLEKIRREAAQMVQYQETQQAHTOQLELLQKFTNKALYSMMRGKL 2289
 Db 2520 GQICVNVLDL-----MTAKGKTTQTNSSSATAPTNETLEVSANFPA-----FLGTI 2566
 Qy 2290 SAIYQFDELQSFCLMAQEAALRELTONGVTIFRGAWNGTTAGIMAGETILLINLAEME 2349
 Db 2567 KA-----NGLVDFSKVLQNTTIGTLDLGNATPK----- 2595
 Qy 2350 KVLDERDERALEVTRTVSLAQFYQALSSDNFNLTEKLFQFLREGKNGVAGSNE-----L 2404
 Db 2596 -----ANLIVNNAFNNNSYRANISGNFVAKGATFTNENGLNVGGNFSEGPLIF 2648
 Qy 2405 KLSNRQIBASVRLSLKIFSDYPES--LGNTRQLKQVSVTL--PALVGPYEDIRAVLNY 2459
 Db 2649 NLNPNHTQIINVTGTSTIMSNQALINFNTQLKQAYTTLINARNVYGV--DNQTL-- 2705
 Qy 2460 GGSIVMPGCCAIALSHGVNDSGQFMDPNDSRYLPPFEGISVNDSGSLTLSPFD 2513
 Db 2706 GGSL-----SDYLLKLYTLIDFNGKRMQLNGDSLSYDNPQVSIKD--GGLVWSFKD 2753

RESULT 8

US-11-051-453-42
 ; Sequence 42, Application US/11051453
 ; Publication No. US20050287150A1
 ; GENERAL INFORMATION:
 ; APPLICANT: AMBROSINO, DONNA
 ; APPLICANT: BABCOCK, GREGORY J.
 ; APPLICANT: BROERING, THERESA
 ; APPLICANT: GRAZIANO, ROBERT
 ; APPLICANT: HERNANDEZ, HECTOR JAVIER
 ; APPLICANT: LOWY, ISRAEL
 ; APPLICANT: MANDELL, ROBERT
 ; APPLICANT: MOLRINE, DEBORAH
 ; APPLICANT: THOMAS, JR., WILLIAM D.
 ; APPLICANT: ZHANG, HUI-FEN
 ; TITLE OF INVENTION: ANTIBODIES AGAINST CLOSTRIDIUM DIFFICILE TOXINS AND
 ; TITLE OF INVENTION: USES THEREOF
 ; FILE REFERENCE: MJI-001
 ; CURRENT APPLICATION NUMBER: US/11/051,453
 ; CURRENT FILING DATE: 2005-02-04
 ; PRIOR APPLICATION NUMBER: 60/542,357
 ; PRIOR FILING DATE: 2004-02-06

; PRIOR APPLICATION NUMBER: 60/613,854
 ; PRIOR FILING DATE: 2004-09-28
 ; NUMBER OF SEQ ID NOS: 82
 ; SOFTWARE: PatentIn Ver. 3.3
 ; SEQ ID NO 42
 ; LENGTH: 2367
 ; TYPE: PRT
 ; ORGANISM: Clostridium difficile
 US-11-051-453-42

Query Match 1.5%; Score 201.5; DB 7; Length 2367;
 Best Local Similarity 19.1%; Pred. No. 0.0002;
 Matches 382; Conservative 324; Mismatches 704; Indels 591; Gaps 110;

Qy 72 NPQLSGAIRLGIERDSRSYDEMFGA-----RSSSEFV-KPGSVASM-FSPAGYLT----- 120
 Db 412 NDTLGPILSQGNDFTTNNFSGESLGAIAENISFIAKIGSYLRVGFPEANTITLGS 471
 Qy 121 -ELYREA-KDLHFSSAYHLNRRPDLADLTLSQSNMDTEISTLSN-ELLLEHITRKT 177
 Db 472 PTIYAGAYKOL-----LTFKMSIDTSLSELRLNFEPPKVNISOAT 513
 Qy 178 GGDSDALMESLSTYRQALDTPYHOPY-----ETIRQVIMTHDSTLSALSRLNPEV 226
 Db 514 EOEKNLSWQFNREAKIQPEYKKNYPEGALGEDNDLDFSQNTVTDKEYLLEKISSSTK- 572
 Qy 227 MQQAEGLASLLAILANISPELYNLTETETKNDALFAQNFSENIITPENFASQSWIAKY 286
 Db 573 --SSEGGVHVIV-----QLOGDKISYEAACNLFAKNPYDSILFQNIEDSEVAYY 622
 Qy 287 G---LELSEVQY-----LGMLOQNGYSD-STSAIY-----VDNIST-----GL 319
 Db 623 NPTDSEIQEIDKYIPDRISDRPKIKLTFIGHGKAENFTDIFAGLDVDSLSEIETAGL 682
 Qy 320 V-----VNNESKLEAYKITRVKTDYDKNINVFDMYEGN--- 354
 Db 683 AKEDISPKSIBINLLGCMFYSVNVETPYGKLLLRVK-----DKVSELMFSMSQDSIIV 738
 Qy 355 --NOFFIRANPKVREF-----GATLRKNAGPSGVISLSPGLIANTNFNKSNYLSNIDSE 408
 Db 739 SANQVEVRLNSEGRELDDHSGEWINKB--ESIUKDISKEYISFNPKENKIIVKSKNL 795
 Qy 409 YKNGVKIYAVRYTSTSATNCGGIFTPEYPLTIFALKLNKAIL--CLTSGLSPLNQL 466
 Db 796 PELSTLLQEIERNNSNS-----DIELEBKVMLEACEINVISNIETQ 836
 Qy 467 TI-VRSNNAQGIINDSV-LTKVFTLYFVSHRYAL-----SFDDAQLNGSVINOYADD 517
 Db 837 VVEERIEAKSLTSDSINYIKNEFKLIESIHALCDLKQONELEDSHFISFEDISETEG 896
 Qy 518 DSVSHFNRLFTPLKGRKIFEADGNVTSIDPDEQSTFARSALMRGLGVNSGELYQLGKL 577
 Db 897 FSIRPINK-----ETGESIFV--ETEKTFSEYA-----NHITEISKI 933
 Qy 578 AG-VLDAQNTITLSVFTVSSLYRLTLARVOL--TVNELCMYGLSPNG--KTASLSS 633
 Db 934 KGTIFDTVN-----GKLVKVKNLDTTHEVNTLNAAFQISLIENYSSKESLSLSV 984
 Qy 634 GELPLVIVLYQ-----VTQWLTEAEITV-----RAIWLCTPERSGNISPEISNLN--N 682
 Db 985 AMKVQVYAQLFTGLNTITDAKVELVSTALDETIDLL--PTLSEGI--PIATIIDGVS 1041
 Qy 683 LRPSISEDMAQSHNRELQAEILAPFIAATLHLASPDMARYILLWTDNLRL-PGGLDI----- 737
 Db 1042 LGAAIKE--LSETSDPLLRQEIETAKIGMAVNLTTATTA-----IITSSLGASGFSILLVP 1096
 Qy 738 -----AGPMTLVKESLNANETTLQV-FCHVMAQLSLSVQTL-----RLSEAEISLVVI 786
 Db 1097 LAGISAGIPSLVNNELVLRDKATKVDFKHVSLVETEGVFTLLDDKVMQOQDDIVISEI 1156
 Qy 787 ---SGFAVIGA-----KNQPAQGH-----NIDTLFS-----LYRFHOWINGLGNPGSDTLML 831
 Db 1157 DFNNSNIVLGKCEIWRMEGGSGGHTVTDDIDHFFSAPSIYR-----BPHLSIYDVL 1207

```
Qy 832 RQQTTLTADRLASVMGLDISMVTQAMVSAGVNQLOCMODINTVLOWIDVASALHTMPSVIR 891
Db 1208 EVQKEELD-----LSKDLMLPNA-----PNRVANE-----TGW-----TPG-LR 1242

Qy 892 TLVNIRYVYALNKAESNLPSPDEWQTLAENMEAGLSTQQAQ----- 932
Db 1243 SLEN-DGFKLDRIRDNVEGEFYRYFAFIADALITLKPRYEDTNRINLDSNTRSFIV 1301

Qy 933 --TLADYABRLSLVLCNWFANLQ-----EGVSLH-SRDDLYSYFLIDNOV-----S 978
Db 1302 PIITTEYIREKLSYFYSGGTYALPLSQYNGMINIELSESDW-IIDVNVVRDVITIES 1360

Qy 979 SAIKTTILAEAI-AGIQLYNR-ALNRTEPNARADVSTRQPFDTWNNRYSTWGWYS-- 1034
Db 1361 DKIKKGLIEGILSTLSEENKIILNSHEINFSGEVNGSGFVSLT-----FSILEGINAI 1416

Qy 1035 -----RLVYPENYIDPTQRIGQTRMDELLENISQSKLSRDTVEDA 1076
Db 1417 IEVDLLSKSYKLLISGELKILMLNSHIQ--QKIDYIGFNSLQKNIPYSFVDSSEKENG 1474

Qy 1077 FKTYLTR-----FETVADLKUVS-AYHDNVNSNTGLTFVQGTRENLPYIYWRNVDISRMQ 1131
Db 1475 FINGSTREGLFVSELPDVVLISKVYMDSDKPSFG-----YYSNNL----- 1514

Qy 1132 AGELAAANAKWETKIDTAVNPYKDAIRPVIFRERLHLI--WVEKEEVAKNQTDPFVETYDR 1189
Db 1515 -----KDV--KVITKQNVNLTGYLKKDDIKISLS--LTLODE 1548

Qy 1190 FTLKLAFLRHDSWSPWSYDITTTQVEAVTDKPDTERLALAAAGFOCEDTLLVYVYKTG 1249
Db 1549 KTIKLSVHLDESVA-----EILKFNKRGSTNTSDSLMSFLSESMNIKSIFVNFLO 1600

Qy 1250 KSYSDFGSNKNVAGMTYGDGSKFKKMENTALSYSQKNFTDIHT--QGN--DLVRKA 1305
Db 1601 SNIKFILDANFIISGTTISIGOFEEICDENNNIOPYIFIKFNTLETNYLYVNGRNQMVIEP 1660

Qy 1306 SY-----REAPQD--FEVPASLN--NGSAIGDSDSLVW-----ENGNIPOI-- 1341
Db 1661 NYDLDDSGDISSTVINFSQKLYGIDSCNVKVISPNYIDETNITPVYETNNYTEVIV 1720

Qy 1342 -TSKYSNDNLAI TLHNAFTVRY-----DGSGNVI-----RNKQISAMKLTGV----- 1383
Db 1721 LDANYINEKINVNIND--LSIRYVNSDNGDNDFILMSTSEENKVSQVKIRFVNVFKDKTLA 1778

Qy 1384 -----DGKSQYGNAFIANTVKHYG-----GYSDLGCGPITVYKNTYNIASVQGHLMN 1431
Db 1779 NKLSFNFSKQDVPVSEIILSFTPSYEDGLIGY-DL-GLVSLYNE-KFYINNFGMMVSG 1835

Qy 1432 ADYTRRLILTPVENNYIARLFEPPFP--NTILANTVFTVGSNKTSDFKKCSYAVD---G 1485
Db 1836 LIYIN-----DSLTY-----FKPPVNNLITGFTVTVGDK-----YYFNPIGG 1873

Qy 1486 NNSQGFQIFSS-----YQSSGWLDDITDGINNTDKITVMAGSKTHTFTASDHIALPANSE 1541
Db 1874 AASIGETIIDDKNYFNSQSVL--QTGVFSEDEGKYFA-----PANTL 1915

Qy 1542 D-----AMPYTFKPLBEIDASSLAFTNPIAIDIVFETKAKDGRV-----LGKIKOTLS 1589
Db 1916 DENLEGAIDPTGK-LIIDENIYIFEDNYRG--AVEWKELDGEHMYFSPETGK-----A 1966

Qy 1590 VKRYNYPEDILFLRETHSGAQYMQLVYRIRLNT--LLASQLVSRANTGIDTILTWETQ 1647
Db 1967 FKGLNQIGDDKYF-----NSDGVNQKGFVSINDNKNHFDGSDGKVMKVGYTEID----- 2014

Qy 1648 RLPEPPLGEGFPANFVLPKYDPAEHGDE-----RWFKIHIHNGVGNTOGRQPYYS 1696
Db 2015 -----GKHIFY-----AENGEMQIGVENTEDGFKYFAHNNEDLGNEEGEBISYS 2058

Qy 1697 GMLSDTSETMTLFPVYABGYTHGVRGLGVGYOKIYDNTWESAFYFDE-TKQOQV-- 1753
Db 2059 GILNFNNKI-----YYFDSDFTAVVGVWGDLE-----DGSKYFYFDDTAEAVIGL 2102
```

```
Qy 1754 -LINDADH---DSCWTOQGIV 1770
Db 2103 SLINDGOYYFNDNDGIMQGVFV 2123

RESULT 9
US-10-995-561-776
; Sequence 776, Application US/10995561
; Publication No. US20050272054A1
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele et al.
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: CARDIOVASCULAR DISORDERS AND DRUG RESPONSE, METHODS OF
; TITLE OF INVENTION: DETECTION AND USES THEREOF
; FILE REFERENCE: CL001559
; CURRENT APPLICATION NUMBER: US/10/995,561
; CURRENT FILING DATE: 2004-11-24
; NUMBER OF SEQ ID NOS: 85702
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 776
; LENGTH: 5935
; TYPE: PR1
; ORGANISM: Homo sapiens
US-10-995-561-776
```

```
Query Match 1.5%; Score 192; DB 6; Length 5935;
Best Local Similarity 17.7%; Pred. No. 0.0042;
Matches 485; Conservative 369; Mismatches 914; Indels 966; Gaps 123;

Qy 51 HETIEQKKNRL-----LEARIFTRANPQLS--GAIRLGIERSVRSYDEMFGRSSS 102
Db 410 HSYINVQNGORLLLDLXELMETLTSRDEYQTSPPKVVVEIGHQROKTPB-----GLQESA 463

Qy 103 FVRPGSVASMPSPAGYITELYREAKDLHPSSAYHLNRRPDLADLTL---SQSNMDETEI 159
Db 464 NVK---ISGTFS-SGMTVRL---PEQFSSQ-----NKEYPDREDCTTEKGGKTTVETED 511

Qy 160 STL-TLSNELLLEHITKTKGSDSALME--SLSTYROAIDTPYHQPYETIRQVIMTHDST 216
Db 512 SSVENPQDILFEVQKERNPIDALKVINKVLEFVORQLIGT-----QREDQT 558

Qy 217 LSALSRNPE-----VMQAEAGASLAILANISPELYNLT-----E 252
Db 559 AVSVRENASRGHLITIPPAEAGVPLV-----VDKOVFSEVETPKKEHQLRNTSFTQNE 613

Qy 253 EITEKNADALPAQNFSENITPENPASQSWIAKYGLEL-----SEVOKYL 297
Db 614 QAHTLETEYTHDETGSGSHIKPQSKLQVQVKTLGIKLEKSETDGNVHPIDKKEMLKKT 673

Qy 298 GMLQNGYSDSTSAYVDNISITGLVNNESKLEAYKITRVTDDYDKNINIFOLMYEGNNQF 357
Db 674 FLAKDHKEQEA--QNIAGSMMWSE-----KTDEED-----SGREI 709

Qy 358 FIRANFKVSRFEGATLRKNAGPSGIVGSLGPIANTFNKSNLYLSNISDSYKNGVKIYA 417
Db 710 FL-----SCSHPLELLEEATLNLV----- 728

Qy 418 YRYTSSTSATNOGGIITFFESYPLTIFALKLNKAILRLTSLGSLPNELOTVIRSDNAQGI 477
Db 729 -----SAQLDGGIF---HEQTGQKLLINEAI-----SRGI 756

Qy 478 INDSVLTKVFTTLPYSHRYALSFDDAQVLNG-----SVINOVADDSDSVHFNLFNTPPLK 533
Db 757 VPSHTAVKMEKL-----NMQGFDSQTCESLITTEEVINEGLMDEKLLH-NVLNADKAIS 811

Qy 534 GKIFEADGNTVSDPDDEQSTFARSALMRGLGVNSGELYQLGK-----LAGVL 581
Db 812 G-----VLDPTOTLCSVKDAVTGVL-----LDKETATRIILERQVVTGGII 852

Qy 582 DAQNTIITLSYFVFISSLYRLTL-----ABVHQLTVNELCMLYGLSPNGKTT 628
Db 853 DLKRGKGVSVTLASTLGLVDVDAQPELINLEKASKGRDAEKTVRE-----RLI SLQMETT 907
```



```
QY 1734 YDNTWESAFFYDEYKQFVLIN-DADH-----DSGWTQOGIVKNYKK-----YKGFANV 1782
DB 1909 GGS-----LDNYLKLALYIDINGKMWMTDNGLTNGQAVSVKDGGLVVGFKDSQNK 1961
QY 1783 STATGY-----SAPMDNSASALYYWELFYTPMCMCFQL--LQEKQFDEATOWI 1830
DB 1962 VIYTSILNKKVIAVNDPINNPQAPTJLQ-----YIAQIQGVQSVDSIDQAGGNQAINWL 2017
QY 1831 NYVNPAGYIVNGEIAPIWNCRLPEETTSWNANPLDAIDPDVAQNDPMHYKIATFMR 1890
DB 2018 NKIPETG--SPLFAPYPLESHSTKOLT----- 2044
QY 1891 LDQLILRDMAYRELTDRALNEAKWYRTLELLGDEPEDYGSQWQAAPSLSGAASQTVQ 1950
DB 2045 -----IAGDIA-----NTLEVI-----ANPNFKNDATNLIQ 2070
QY 1951 A-YOODJTLMLRGVGSKNLTANSVLGLFLPEYNPALTDYMWTLRLRLNLRHMLSDG 2009
DB 2071 INTYTOQMSRLAK-----LSDTSTF-----ARSDFL-----RLEALKRFRADA 2110
QY 2010 OPLSLAIYAEPTDPRKALLTSMVQASQGSVALP-GTLSLYRFPVPMLE-TRNLV----- 2061
DB 2111 IPNAMDVLKYSQRNRVKNVWATVGGASFSIGGTGLYGINVGYDRFIKGVIVGGYAA 2170
QY 2062 -----AQLTOFGTSLLSMAEHDDA-----DELTTLLLOQG-----MELATQ 2097
DB 2171 YGSGFHANITQSGSSNVVGVYSRAFIKRSLSLNETWGTYNKTFINSYDPLLSIINQ 2230
QY 2098 SIRIQORTVD 2107
DB 2231 SYRYDTWTTD 2240

RESULT 11
US-11-013-759-9
; Sequence 9, Application US/11013759
; Publication No. US20050249747A1
; GENERAL INFORMATION:
; APPLICANT: Loomore, Sheena M.
; APPLICANT: Sasaki, Ken
; APPLICANT: Yang, Yan Ping
; APPLICANT: Klein, Michel H.
; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
; TITLE OF INVENTION: PROTEIN OF MORAXELLA
; FILE REFERENCE: 1038-921MIS:1b
; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 2053
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-9

Query Match 1.4%; Score 182; DB 7; Length 2053;
Best Local Similarity 18.3%; Pred. No. 0.0031;
Matches 339; Conservative 221; Mismatches 609; Indels 680; Gaps 86;

QY 110 ASMFSPAG-YLTLEYR-----EAKLHFPSSAYHLNRRRLADLTLSQSNMDE---IS 160
DB 293 ASAVVPLGKTLADQYKATROGSDTDI-FSIGNNNNNSSIRRKINVAGSRDIDAVNVA 351
QY 161 TLTLNELLLEHITKTDGSDALMESLSTRQALDTPHQPYETIROVIMTHUSTLSAL 220
DB 352 QLKVLELANRKITFKGDDNN-----SNSVERGLGT-----LTIKGDAQTNAL 396
QY 221 SR-NPEVNGVQAG-----ASLLAILANISPELYNLTETETKKNADALFAQNFSENITPE 274
DB 397 TEANIGVVDGNGLVKLAKELTGLTSVS-----ATNKITVSNWNNNABLSQSGLTFS 450
```

```
QY 275 NFASQSWIAKY--GLELSE-----VQKYLGLM-QNGYSDSTSAYVDN---- 314
DB 451 PITGKTDTKYSIDGLAFTINDSNSIATGTRITTKKIGFAGTGVGDVDESKPYLDNEKL 510
QY 315 -----ISTG-LVNN-----ESKLEAYKITRVK----- 336
DB 511 KVGSTLSSGLTWNNTTGNKQIQVANGIKPATVANVANTSATVGTARITEEIKGFAG 570
QY 337 -TDDYDKNINFDLMYEGNNOFFIRANPKVSR-EFGATLRKNAGPSGIVSGLSGLPIANT 394
DB 571 TNDGVEQAPYLD-----KERLKVGRVEITTDGGINAGNHKITGLTNG--IANT 617
QY 395 NFKSNYLSNISDSE-----YKNGVKIYARYTSSSATNQGGGIFTFESYPLTIPALKLNK 450
DB 618 DAVT--IKQLDKAKPTLNAGDGI--NSNNGDLVDSSGNITPTTYNITSVKTTLN- 669
QY 451 AIRCLTSLGSLPNELOTVIRSDNAQGIINDSVLTGVFTLFFVSHRYALSFDQAOVLNGSV 510
DB 670 -----SNGTSGNNKFSV--SNAHD--NNSLVT-----AKLDADYLN--K 702
QY 511 INQVADDDSVSHFNRLFNTPPLKGIKIFEADGNVTSIDPDEQSTFARSALMERGLGVNSGE 570
DB 703 VNETADS-----ALPSFKVQNGDGNNAITVGKDTNGKTFNTLKLKGENGVN--- 749
QY 571 LYQLKLAGVLDQANTITLSVFVSSLYRLTLARVHOLTVNEL--CMYGLSPFNKTKT 628
DB 750 -----ITNRAATGTVTFGIDQSNGLTT 771
QY 629 ASLSSGELP--RLVIMLYQVTQMLTEAEITTEAILWLCTPEFSGNISPEISNLNLRP 685
DB 772 PKLTVGSDTNGNRLVI-----EQV-----PSADGN---STYKNIIGLSP 807
QY 686 SISEDMAQSHNRELQAEILAPPIAATLHASPDMARYILLMTDNLRLPGLDITAGFMTLVL 745
DB 808 TL-----PSIAS-----PSGRNIALGNTIEE 828
QY 746 KESLNANETTOLVQFCHVMAQLSLSVQTLRLSEAELSVLVSGFAVLKAKNQAPQACHID 805
DB 829 KDKSNA-----ASIDDLNAGP---NLKNGKDKDFVS 858
QY 806 TLFSLYRPHOWINGLNGPSDTLMLROOT-----LTADRLASVMGLDISM 851
DB 859 T-----YDTVDFIGNATTATVYDEANQTSKAVADVNVDEKTIETLTGNGKKQLGVKTIK 914
QY 852 VTQAMVSAGVNLQOCWQDINTVLQWIDVASALHTMPSVIRT-----LVNIYVYA 901
DB 915 LTETSTNGNATFTST-DDDHVAVKASDIAGNLNTLAEIHTTKGTANTALQTFVKKVDE 973
QY 902 LNKAESN-----LPSWDEWQTLAENMEAGL----- 926
DB 974 NDKADDTNATVVGKDTSGKVNTLKLKGNGLDIDTKDKGTVTFGINTQSGLKAGDSTTL 1033
QY 927 -----STQOQOTLAD-----YTAERLSSVLCNWL--- 951
DB 1034 NNGLSINKNTASNQIQVGADGVKPMVNVNGVAGIDGTRITRDEIGFTGTSGLSKS 1093
QY 952 -----ANIQPEGVSLHSRDDLYS---YFL---INDVSSSAIKTTRLAEA 989
DB 1094 KPHLSKDGINAGGKKITNIQSGEIAKNSHDAVTGSKIYDLKTELENKISSTAKT----- 1147
QY 990 IAGIOLYINRALNRIEPPNARADVSTRQFTDWTNNRYSTWGVSR---LVYIPENYIDP 1046
DB 1148 -----AQNSLHEFSVADEQGNF---TVSNPYSSYDTSKTSDVITFAGENGTT 1193
QY 1047 TQRIQTQRMDELLENISQSLSRDTVEDAFKTYLTRFETVADLK---VSAYHDNVNSN 1103
DB 1194 KWKGVVRVGIQDTKGLTTPKL-----TVGNNGKGIIVINSQNGQNTI 1236
QY 1104 TGLTFVFGQTRNLPPEYWRNVDIRMOAGELAAANAKEMTKIDTAV--NPYKDAIRPVI 1161
DB 1237 TGLS-----NTLANVTNDKGSVRTTEQGNIIKDE--- 1265
QY 1162 FRERLHLIWEKEBEVAK-----NG--TDPVETYDRP-----TLKLAFLR 1198
```

```
Db 1266 -----DKTRAASIVDVL SAGFN LQNGEAVDFVSTYD TVNPFANGNTTTAKVTY-- 1313
Qy 1199 HDGSKSAPWSYDIT--TQVEAVTDKDPDTERLALAAAGFGOGEDTLLVVFVYKTKSY-- 1252
Db 1314 DDTSKTSKVYVDVNVDDTTIE-VKDKKLGKVTTLTSTG-TGANKFALSNOATGDALVKA 1371
Qy 1253 SDF-----GGSKNVAGWTIYGDGSKFKKMENTALSRYLSOLKN--TFD--- 1292
Db 1372 SDIVAHNLTSGLDQTAKGASQANNSAGYVDADGN-KVIYDSTONKYFOAKNDGTVDKTK 1430
Qy 1293 -----IHTQGNLVLVRKASYR-----FAQDFEVPASLN-- 1320
Db 1431 EVAKDKLVAQAQTPDGTTLAQNVRKSVINKEQVNDANKKQGINEDNAFVKGLEKAASDNKT 1490
Qy 1321 MGSAL-----GDSLTVWNGNIPOI-----TSKYSSDMALITLHNAA 1358
Db 1491 KNAAVTVGDLNVAQAQTPLTFAAGDTGTTAKKLGETLTITKGQTDNKLTDNNIGVAGTDG 1550
Qy 1359 FTVR-----YDGSNVIRNKQLSAMKLTGVDGKSOXGNAPFIANTVKHYGGYSDL 1408
Db 1551 FTVKAKDLTNLNSVAGGTIKDEKIS---FVDANGQAKANTPVLGANGL-----DL 1600
Qy 1409 GGPITVYNTKKNYIASVQGHLMNADYTRRLILTPVEN--NYIARLFFPPFSPN----- 1459
Db 1601 GGVV-INSVNGKTKDQDAANVQQLNEVNLGLGNDNADGNQVNIADIKKDPNPGSSSNR 1659
Qy 1460 TILNTVTVGNSKTSDFPKCSYAVDGNNSQGFQFSSYQSGGWLDDIDTGIN-----NTDIK 1515
Db 1660 TVIRAGTVLGGKGNNDTEKLA-----TGGVQV-----GVDKDNANGDLSNVWVK 1704
Qy 1516 ITVMAGSK---THFTFA---SDHLASLPANSPFDMPTFFKLELDASSLAP--TNNTIAPL 1567
Db 1705 -TQDGGSKKALLATYNAGQTNVYNNPAAEID-----RINEQIRFFHVNDGNQE 1754
Qy 1568 DIVPETRAKDGRLVKIKQTLISVRVNVNPNPDILFLRETHSGAQMOLG 1616
Db 1755 PVVQGRNGIDSASSGKHSVAIGFOAKADGEAAVAIGRQTQAGNSIAG 1803

RESULT 12
US-11-102-476-46
; Sequence 46, Application US/11102476
; Publication No. US20050271680A1
; GENERAL INFORMATION:
; APPLICANT: Chang, Yung-Fu
; APPLICANT: Palaniappan, Raghavan U.M.
; APPLICANT: Cornell Research Foundation, Inc.
; TITLE OF INVENTION: Novel Immunogenic Proteins of Leptospira
; FILE REFERENCE: 1153.080US1
; CURRENT APPLICATION NUMBER: US/11/102,476
; CURRENT FILING DATE: 2005-04-08
; PRIOR APPLICATION NUMBER: PCT/US2003/32385
; PRIOR FILING DATE: 2003-10-10
; PRIOR APPLICATION NUMBER: US 60/417721
; PRIOR FILING DATE: 2002-10-10
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 46
; LENGTH: 1889
; TYPE: PRT
; ORGANISM: Leptospira interrogans
US-11-102-476-46

Query Match 1.4%; Score 178.5; DB 7; Length 1889;
Best Local Similarity 18.8%; Pred. No. 0.0046;
Matches 377; Conservative 255; Mismatches 671; Indels 701; Gaps 95;

Qy 7 LNKISPTRDQQTWTLADQLVLSSEL-----RKIPDQLSWGEARLHYHETIQKKNR 61
Db 226 VLSQIQTNNLNPGLGKKQKLTATGIYSDNSNRDSSVWNSS-----NSTIANIQNG 281
Qy 62 LLEAIRIFRANPQLSGAIRLGIERSVRSYDEMFGARSSSFVKPGSVASN-FSPAGYLT 120
```

```
Db 282 ILETAD-----TGIV-----TVSASTENIIGS-VKLVITPAALVLSVSPSTNSTV 325
Qy 121 ELYREAKDL--HFSSSAYHLDNRPPDLAD-LTLSQSNMDT-----EISTLTLSN 166
Db 326 -----AKGLOENFKATGIFTDNSNSDITDQVTDWSSNTDILSISNASDSHGHLASTLNOGN 380
Qy 167 ELLLEHITRKTGG-----DSDALMESL-----STYRQAIPTFPYHQPYETI-- 206
Db 381 ----VKVTASIGGLOGSTDFKVTOEVLTSIEVSPTRTSIAKGLTKQKTAIGTIFDTSKKD 436
Qy 207 --RQVIMTHDSTLSALSARNPE-----VMQOAGASLL-----AILANI- 242
Db 437 ITDQVTWNSSSAIVSVSNLDDNKGKGAHAVGDDTTITATLGKAVAGKTWLTVPVPAVLSIQ 496
Qy 243 ----SPELYMLITEETEKADALFAONFSENITPEN--PASQSWIAKYVLEISEVQKY 296
Db 497 INPVNPSLAKGLTKQFT---ATGIYSDNSNKRDIITSAVTFWSSDSSIA-----TISNAQKN 548
Qy 297 LGMQLQNGYSSTSAVDNIISTGLVNNESKL-----EAYKI 332
Db 549 QG---NAYGAATGTTDIKATFGKVSSPVSTLSVTAAKLVEIQITPAAASKAKGLTERPKA 605
Qy 333 TRVKTDYDKNINYFDLMYEGNNQFFIRANPKVSRFEGATLRKNAGPSGVISLGSGLPIA 392
Db 606 TGIFTDNSNDI-----TNQ-----VTWNSSNTDIAEIKNTSGSGKITNTLT---P 648
Qy 393 NTNPKSNYLNISDSEYKNGVKIYAYRYTSTSATNOGGGIFTFESPLTIFALKLNKA- 451
Db 649 GSSEISAALGSIKSSK-----VILKVTTPAQ 673
Qy 452 -IRLCLTSGLSPELQITVRSDNAQGIINDSVLTKFVTLFYSHRYALSFDQDAVLN--G 508
Db 674 LISIAVTP-INPSVAKGLIQFRATGYTDHVSQDVTALATWS---SSNPKAMVNVVTG 729
Qy 509 SV-----INQVADDSDSVSHFRLFNTPPLKGIKIFEADGNTVSIIDPEEQSTFARSALM 561
Db 730 SVTTVATGNTNIKATIDISIGSSSVLNVTPALL-----TSIBITPTINSITHGLTKQF 781
Qy 562 RGLGVNSGELYQLGKLAGVLDQAQNTITLSVVFVSISSLYRLTLARVHQLTVNELCMLVGLS 621
Db 782 KATGIFSDK-----STQNLTLQTLVTWISSDPSKIE-----IENTS 815
Qy 622 PFNGKTTAS-LSSGELPRLVILWYQVTOQLTEABITTEAIWLLCTPFSGNISPEISNLL 680
Db 816 GKGIATASKLGSNNIKAVYKVFQSSPIPIVTDLKLKSI-----TISSSSSIA 865
Qy 681 NNLRPSISEDMAOSHNRLEQAEILAPFIAATLHLASPDMARYILLWTDNLRPGGLDIAGF 740
Db 866 KGL-----TQQKAIGTFIDG- 881
Qy 741 MTLVLKESLANETTQLVQFCHVMAQLSLSVQTLRLSEABLVLVSGFAVLGAKNQAPAG 800
Db 882 -----SEQETNLV-----TWYSSKSDVAPINNAAKGLATALSIG 918
Qy 801 QHNIDTLFLSYRPHOWINGLNGPCSDTLDMLROQTLTADRLASVMGLDISMVTQAMVSAG 860
Db 919 SSNISAIVNSISNNK-IN--FNVSAATLDSIKINPVN-NNIAK--GL-----TQOYTALG 967
Qy 861 VNQLQCMQDINTVLQWIDVASALHTMPSVIRTLVNIIRYVTALNKAESNLPSWDEWQTLAE 920
Db 968 VYSDTTIQDISDSVTM---SSSNSSSISISNSTETKGKATALQIGKSKITA--TYSNISE 1022
Qy 921 NMEAGLSTQQAQTLADYTAERL-----SSVLNWFANIQPEGVSLHSRDDLYSYFL--- 972
Db 1023 NIDI---TVSAAATLSSISISIPINTINATVSKQPFAMGTYSVG---TKADLTSSVTWSS 1075
Qy 973 ---IDNVSAAIKTRTAAEAIA-----GIOLYINRLNRINPNARADVS- 1013
Db 1076 SNKSQSKVSNARKTKGLVTGIASGNSIITATYGSVSGNTILTYNKT-DTAPTQSVSVSL 1134
Qy 1014 ---TRQPFDTWTVNNRYSTWGGVSRL-VYYPENYI-----DPTORIGOTRMMDELLENISQ 1065
```

Db 1135 SPPTIQVYVESINNKEAL--DLSNYKIINSNFGHCSNDTDFNSNSQTADFSLSIKG 1192
QY 1066 SK-----LSRDTVDAPKTYLTRPETVADLKVVSAYHDNNSNTGLTFWVGQTRNLPE 1119
Db 1193 SKNFTIYLSHQILINKSYTLVVKQGHDLSSI-----PNSLSCPNNSDFIG----- 1240
QY 1120 YYWRNVDISRMOAGELAANAKWKTKIDTAV-NPYKDAIRPVI FRERHLHIWKEBEVAK 1178
Db 1241 -----KEQLKLTSAVCNSLNQVI--VSFSKPLY-----SGKEATK 1273
QY 1179 N--GTDPVETVDRFTLKLAFLRHDSWSAPWSYDITTOVEAVTDK-----KPDTERLALA 1231
Db 1274 SVECSNPSQCESRY--KEAGVSSLSG-----ITSVRILDGKVGCGAPADSSKICLT 1322
QY 1232 ASGFGEDTLLVYVYKTSYDFGGSKNKNAVGMTIYGDGSKKMENTALSRYSLKNTF 1291
Db 1323 HSLLOS-----GGQYIIAANDLNGDGNK-----SWGAIROSF 1357
QY 1292 DIIHQ-----GNDLVKASY----- 1307
Db 1358 DQENLQSPKDKRINEFGCNSPLNFMGPIVSDPFGDGSDFGLVDYNNQIYLGPNVKN 1417
QY 1308 ---RPAQDFEVPASL--NWGSAIGDSDLTVMWNGNIPOITSKYSDNLAIILHNAFTV- 1361
Db 1418 QAARFNYDGTTPESIFFSTQDINATNRASRDGGIP-VPNYVTIHTGCTILNSADITG 1476
QY 1362 ---RYDGG-----NVIRNKQISAMKLTGVD 1384
Db 1477 CGPNEDGRGVFATGSLDKKSHIFTAGSKPKSFNYLYSSDTHNLNFKYISMGKITGL- 1535
QY 1385 GKSQYGNAFIIANTVKHYGGYSDLGPIITVYNKTKRYIASVQHLM--NADYTRRLILTP 1442
Db 1536 --ATAGTSSIAVLDDRIHVGA-----KKNQNLNAPDPGKITFNTSEHNRCAIVNN 1584
QY 1443 VE-----NNYVABLPFPP-----SPNTILNTVFTVGSNKTSDFKKCS 1480
Db 1585 CEASDGYRGNRFP--RIDRMPYFGGSVDVWVYRSYKSDNSSINWGYVYVIGDLSLFVFKSKL 1642
QY 1481 YAVDG--NNSQGFQIFSSYOS-----SGWLDITGINNTDIKITVMAGSKTHT 1526
Db 1643 YAANGFPNSLNGSIHISTSANPSPCEGINRCSWKDTAPRSN-----PKWH- 1690
QY 1527 FTASDHIALSPANDAMPYTPFKPLEIDASSIAFTNNIAPLDIVPETKAK-DGRVLGKIK 1585
Db 1691 -----NSPHNWFSL-----ELTKYRNLI PADKAPSPAEFNGRL----- 1725
QY 1586 QTLVSRVNYNPEDILFRETSHGAQYMQLGVRIRLNTLASQLVSANTGIDTILTWE 1645
Db 1726 -----YVTRTICVTKEDHSGLRQ-----SLQTVKGTG-----SY 1756
QY 1646 TORLPEPPLGEGFFANFVLPKYDPAEHGD-----ERWFKIHIGNVGN-TGRQPYSGM 1698
Db 1757 TNRFPQ-----LWKCDPILTGDTTCEADW-----SLVGNGTGTNFP----- 1795
QY 1699 LSDTSETSWTLFPVPAEYGMHEG 1722
Db 1796 -GDSNHSMTMV--ASGSYLVI 1816

RESULT 13

US-11-075-185-28
; Sequence 28, Application US/11075185
; Publication No. US200502664341
; GENERAL INFORMATION:
; APPLICANT: REEVES, CHRISTOPHER D
; APPLICANT: JULIEN, BRYAN
; APPLICANT: REID, RALPH
; TITLE OF INVENTION: BIOSYNTHETIC GENE CLUSTER FOR AMBRUTICINS
; FILE REFERENCE: 010099.03
; CURRENT APPLICATION NUMBER: US/11/075,185
; CURRENT FILING DATE: 2005-03-07
; PRIOR APPLICATION NUMBER: US 60/551,103
; PRIOR FILING DATE: 2004-03-08

; PRIOR APPLICATION NUMBER: US 60/568,290
; PRIOR FILING DATE: 2004-05-04
; NUMBER OF SEQ ID NOS: 61
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 28
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-11-075-185-28

Query Match 1.3%; Score 170.5; DB 7; Length 2087;

Best Local Similarity 18.2%; Pred. No. 0.019;
Matches 406; Conservative 304; Mismatches 860; Indels 661; Gaps 103;

QY 532 LKGIKFEADGNTVSDPDDEQSTFARSALMRLGLVNS--GELYQGLKAGVLDAQNTITL 589
Db 128 LDGKIADTFNDVEMN--EKFAPELERLSRAVGKEGIGQVRSMGEVSGAW-ADEVASV 183
QY 590 SVFVISSLYRLTLARVHQLTV-----NELCMLYGLSPFNKTTASLSSGELPRLVIWLY 644
Db 184 NALIGDLVQPTREMARVIGAVAKGDLSTQMALEVGRLEGE---FLQTAQTVNTV--Y 238
QY 645 QVTQWLTEA-----EITTE-----AIWLLCTP---EPSGNISPEISLNLNL 683
Db 239 QLGSFASVTRVAREVTEGKLGGOAEVKGAVGTWKDLTDSVNSMAGNLTAQVRNIAVT 298
QY 684 RPSISEDMAQSHNRELQAEIL-----A 705
Db 299 TAVANGDLTKITVDVRGEILELKDFTFMTVDQLRSPASEVTRVAREVTEGKLGQASV 358
QY 706 PFIAATLHASPDMARYILLMTDNLRPGLDIAGFMTLVLKESLNANETTQLV-QFCHVM 764
Db 359 PGVAGTW---KDLTDSVNSMASNLTAQVRNIAAATTAVANGDLTKITVDVKGEILELK 414
QY 765 AQLSLSVQTLRLSABLSVLVSGFAVLGAKNPAGQHNIDTLPFLSRFHQWINGLGNPG 824
Db 415 DTFMTVDQLRSPASEVTRVA---REVTEGKLGQAE-----VKGVAGTW 457
QY 825 SDTLDMLRQQLTLADRLASVNLGLDISMTVQAMVSAGVNLQOCWQDINTVLOWIDVASALH 884
Db 458 KDLTDSV--NSMASNLTAQVR--NIAAVTTAVARGDLTQ-KITVDVRGEI--LELKDTFN 510
QY 885 TMSVIRTLVN-IRYVTAALKAESNLPSWDE-----WQTLAENWEAGLSLQOQATLAD 936
Db 511 TMDQLRSFASVTRVAREVTEGKLGGOAEVKGAVGTWKDLTDSVNSMASNLTVQ-LRD 569
QY 937 YTABRLSVLCNWFEL--NIQPEGVSLHSRDDLYSYFLIDNOVSSAIIKTRLA----- 987
Db 570 --VSKVATAIANGDLTKITVDVRGEILELQIKDVINT--TVQLSFPAEAEVTRVARDVGE 625
QY 988 --EAIAGIQLYINRALRIEPAERADVSTROFTDWTVNNRYSTWGVSRVLY 1038
Db 626 GKLGGOAEVKGAVGTWKDLTDSVNSMASNLTAQ-----VRNIAAVTTAVAR--- 671
QY 1039 YPENVIDPQIRIGQTRMMDELLENISQSLSRDITVEDAPKTYLTFETVADLKVVSAHYD 1098
Db 672 ---GDLTKI-TVDVRGEILE---LKNFTNTMDQLRSFAAQVTRVA----- 711
QY 1099 NVNSNTGLTFVGTQTRNLPYYWRNVDISRMQAGEL---AANAWEKMTKIDTAVNPK 1154
Db 712 -----REVTEGKLGQAEVTVAGVTWKDLT----- 737
QY 1155 DAIRPVI FRERHLHIWKEBEVAKNGTDPVETYDRTLLKLAFLRHDSGWSAPWSYDI--- 1211
Db 738 DSV-----NSMASNLTAQVRNIAADVTTAVA-----NGDLSKKITVDVRGE 777
QY 1212 -----TTQVEAVTDKPDTERLA--LASGFGEDTLLVYVYKTSYSDFGSN-- 1259
Db 778 ILELKDFTNTMDQLRSFASVTRVAREVTEGKLGGOAEVKGAVGTWKDLTDSVNSMAS 837
QY 1260 -----KNVAGWT---IYGDGSKKMENTALSRYSLKNTPTDI IHTQGNLDRKASYPFA 1310
Db 838 NLTAQVRNIAADVTTAVARGDLS--KKITVDVKGEILELKNTP-----NTWVDQLSSFAA 889

Qy 1311 QDFEVPASLNMGSAIGDSSLTWEMNGNIPOITSKYSS--DNLAITLHN-AAFT----- 1360
Db 890 EVTRVAREVGTGKLGQAQAEVTVAGTWDKLTDSVNSMASNLTAQVRNIAAIVTAVANGD 949
Qy 1361 ---VRDGSNGVIRNFIQISAMKLTGVGDKSQYGNAPI--IANTVKGHYGYSIDLGGPITV 1414
Db 950 LSKKITVDVRGEIIELEK--NTINNTWDDQ-----NAPASEVTRVAREVGTGKLGQAQSV 1003
Qy 1415 -----YNTKKNYIASVQGHLMNADYTRRLILITPVNNYYARLFEPPF-----SPN 1459
Db 1004 PGVAGTWDKLTDNVFNAGNLTNVRGIKAVVTVAVANGDLKRLAFDAKGEIAALADTIN 1063
Qy 1460 TILNTVTVGNKTSDFKCKSYAVDGNNSQGFIFSSVQSGWLDIDTGIN----- 1511
Db 1064 GVIEITLAFADQVTVVARE--VGVEG--KLGQAQSVFGAAGTWDKLTDNVFNQLANITQ 1119
Qy 1512 ----TDIKITWAGSKTHTF--TASDHIASLPANSFDAMPYTFKPLFIDASSLAFTNNIA 1565
Db 1120 VRAIAEVATAVKGDLTFTIRVEAQGEVASL--KDTINEMIRNLKDTYLNKSE----- 1170
Qy 1566 PLDITVFTKADGKVLGKIKOTLSVKRVNPNEDILFLRETHSGAQVQMGVYRIRLNTL 1625
Db 1171 -QDWLKTNLAKFSRLQKQDLLTVGR-----L 1197
Qy 1626 LASOLVSRANTGIDITILMTETORLPEPPLGEGFPANFVLPKYDPAEHGDERWPKIHIGNV 1685
Db 1198 ILSLAPVVGAAQGVFFTMVAK--EPTIL--KLLASYA---YKVRKHVDNH--FKLGEGVL 1250
Qy 1686 GGNTGRQPYSGMLSDTSETMTLFPVPAEGYVMHGVRLGVGYOKITYDNTWESAFYF 1745
Db 1351 G-----QCAL--- 1255
Qy 1746 DETKQFVLINDADHDSGMTQOGIVKNIKKYKFLANYSIATGYSAPMDFNSALYIYWE 1805
Db 1256 --EKEKILLVN-APPD-----YIRITSGLGEAPPVNIIVPLFEGQV 1295
Qy 1806 FYTPMFCFQRLLEKQ--PDEATQWVYVNPAGIYVNGEIAPIWNCRPLETTSWNA 1863
Db 1296 KAVIELASFERFSPHQAFLDQLTESIGIVLNTTEANNRTE-----DLKQOSOLA 1346
Qy 1864 NPLDAIDPDAVAQNDPMHYKIATFMRLLDQILRGDMAYREL--TRDALNEAKMYYVRTL 1921
Db 1347 RELSQEELQOTNAELGEK-----ARLLAQNVVEKRGVEQAQALEEK----- 1394
Qy 1922 ELLGDEPEDYGSQWQAAPSLGASQTVQAAIQDQLTWLGRGVGSKMLRT-ANSLVLGLFL 1980
Db 1395 -----ARQLAITSKYKSEFL-----ANMSHELRTPLNSLL----- 1424
Qy 1981 PEYNPALTDYQTLRLRLFNLRHNLSDGQPLSLA--IYAEPTDPKALLTSMVQAS--QG 2036
Db 1425 -----ILSDQ-----LSKNDRNLTGRQVEFAKTIHSGNGDNLALINDILDLSKIES 1471
Qy 2037 GSAVLP--GTLISLRYFPVNLERTNLVAQLTOFGTSL-----LSMAEHDDADELTLLLQ 2090
Db 1472 GTVIDVGEISFSDLODYVERTFRHVAESKLEFELNPAQNLPOVIYTDKRVQOVL--- 1528
Qy 2091 GMELATQSIRIQRTVDEVDADIIVLAESRRSAQRLEKYQOQLYDEBINH-GEORANSL 2149
Db 1529 -KNLSNSFKPTEGSRVALDVD--LVTSGWAPEN-----EGLSRAGAIAIAMSVR 1574
Qy 2150 DAAG-----QSLAGQVLSIAEGVADLPVNVFGLACGSRWGAIRASASVM-----SLS 2199
Db 1575 DTGIGIPHDKQOIIFEAQQADGST-----SRKYGGTGLGLAISREIAWMLGGEIKLS 1627
Qy 2200 ATASQYADKILSRGEAY-----RRRQEWEIFORDNADGEVKQMDAQLESIKIRREAAQOMVE 2256
Db 1628 SKPGSGSFYLYLPLTYTPAPRRKEQTVEVPSAPPVAVSGDVP-----PRSA----- 1676
Qy 2257 YQETOQAHTAQOELLQKFTNKALYSWMRG-KLSAIYVQYFDDLTQSFCLMAQALAREL 2315
Db 1677 -----EPPPHLLNGQVDDSA-----GLKPSDSVVLIVENDASFAHFMVDVAH--- 1718

Qy 2316 TDNGVTFIRGAMNNTTAGLMAGE-----TLILNLAEMEKVW--LERDERALEVTRTVS 2367
Db 1719 -DHGFKAIL--AVRGGAALSIVERRRVAITLIDINLFDMDG--WRVLDREVXKDLB--TRHIP 1773
Qy 2368 LAQFYQALSDNFNLTEKLTQFLRBGKGNVGAAG--NEKLKLSNRQIEASVRLSDLKIFS 2424
Db 1774 V---QVITTD-----EERERALR-----MGAKGVLCPLKTRDALDETFRRLSQFMVSS 1819
Qy 2425 -----DYPESLGNTRQLKQVSVTLPALVGPYEDIRAVLNYGGSIVMPRG 2469
Db 1820 RRKIVLAGPDARQELVELLGGD-----DVTIRSVASGEALDALVTERPDVLLIR-- 1871
Qy 2470 SAIALSHGVNDSQGFMDLDFDNDNSRYLPFEGISVNDSSG-----LTLSPDADTDROKALIES 2524
Db 1872 -----IDLDPVRCFDLIGLAQSGSGTDLPLVLYVAPEEISPADAQLSR 1915
Qy 2525 LSDIIL--HIR 2533
Db 1916 FSQLMVLKHVR 1926
RESULT 14
US-11-113-424-14
; Sequence 14, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 2769
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-14

Query Match 1.3%; Score 167.5; DB 7; Length 2769;
Best Local Similarity 18.1%; Pred. No. 0.048;
Matches 352; Conservative 244; Mismatches 598; Indels 755; Gaps 99;
Qy 826 DTLDMLRQOITLTADRLASVMGLDISMVTQAMVSAGVNLQLOC-----MODINTVLQWIDVA 880
Db 1119 KDTDVYNNQK-----VFGL-----SEAFVSVGVYESCPDLILKEKRTTVLQGYEID 1164
Qy 881 SALHTMPSVIRTLVNIIRYVVTALNKAESNLPFSWDEWQTLAENMEAGL-----STQQ 930
Db 1165 A-----SKJGWSLKKHALNIQSGILHKGNGENQFVSQ 1199
Qy 931 AQTLDYTA-----ERLSSVLCNW-----FLANIQPEG- 958
Db 1200 PPVIGSIGMNGRRRSISCPCSCNGLADGNKLLAPVALTCGSDGSLYGVDFNIRIFPSGN 1259

QY 959 -----VSL-----HSRDDLYSYFLIDNOVSSA-----IKTT-----RLAE 988
Db 1260 VTNILELRKNDPGRSHSPAKHYLLATDPMSCAVFLSDSNRSRVFKIKSTVWVKDLVKNSE 1319
QY 989 ATAGIQLYNRAL-----NRIEPNARADVSTROFFDWTNNRYSTWGGVSLRVYYPENVI 1044
Db 1320 VVAGTG-----DQCLPDDTRCGDGGKA-----TEATLTNPRGTVDKFGLIY-----FV 1364
QY 1045 DPT--QRIGQTRMDDELL--ENISQSKLSRDTVEDAFKTYLTRPETVADLKVWSAYHDN 1099
Db 1365 DGTMRIRIDQNGIISTLLGSNDLTARSPLSCDSV-----1398
QY 1100 VNSNTGLTFWVGQTRNLPEYFWRNVDSRMOAGELAAWAKWETKIDTAVNPYKDAIRP 1159
Db 1399 -----MDISQVRL-----EW-PTDLAINPMDNSL--1421
QY 1160 VIFRERLHIWEKEEVAKNGTDP--VETVDRETL-KLAF-----LHDSGW 1203
Db 1422 YVLDNNVVLQISENHQVRIVAGRPMHCQVPGIDHFLSKVAIHATLESATALAVSHNG--1479
QY 1204 SAPMSYDITTOVEAVTDKPP-----DTERLALAASG-----FGEDTL 1241
Db 1480 -----VLIAETDEKINRIQVTTSGEISLVAGAPSGCDCKNDANCDCFSGDDGY 1530
QY 1242 L-----VFVYTKSY-SDFGG-----SNKNVAGMTIYDGSFKME 1277
Db 1531 AKDAKLTPSLAVCAGDELYVADLGNIRIRIRKNKPLFNTQNNYELSSPIDQELYLF 1590
QY 1278 NTALSRYSOLKNTFDIITQGNDLVRKASYRPAQDFEVPASLNMSAIGDSDLTWMNGN 1337
Db 1591 TGKHLTYQSLPTGDIY-----NFTYGD-----GDITLITDNNGN 1627
QY 1338 IPQITSKYSSDNLAITLHNAFTVRYDGSN--VIRNKQI-----SANKLTGYDG 1385
Db 1628 M-----VNVREDSTGMLVLVDPGVYVWMTGNSALKSVTTQG 1667
QY 1386 -----KSQVGNAPIIANTVKHYG-----GYSDLG-----GPITYNKTKNVIASVQ 1426
Db 1668 HELAMMTYHNSGLLATKSNENGWTTFYEDSFGRLTNVTFTQGVSSFRSDTSVHVQ 1727
QY 1427 GHLMNAD-----YTRLLILTPVNNY-----ARLP-----EPFSPN 1459
Db 1728 VETSKDDVTITNLSAGAFYT--LLOQVNRNSYIIGADGSLRLLLANGMEVALQTEPH 1785
QY 1460 TILNTVF--TVGSKNTSDFKCKSYAVDGNNSQGFQIFSSYQSSGDLIDTGINNTDIK---1515
Db 1786 LLAGTVPNTVGRNVT-----LPIDNGLNVEWRQK 1817
QY 1516 -----ITVMAGSKTHFTASDHIASLPANSPDM-----PYTFKPLEIDASSL 1558
Db 1818 EQARGQVTFV--GRRLRV-----HNRNLLSLDFRVTREKIDYDDHRKFTLRILYDQAGR 1871
QY 1559 AFTNNIAPLDIVFETKAKDGRVLGKIKQTLVKRVNNPNE-----DILFLR 1604
Db 1872 SLWSPSSRLNGVNTYSPGGYTAG--IORGIMSERMEYDQAGRITSRIFADGKTWSTYILE 1930
QY 1605 ET-----HSGAQMOLGVYRIRLNTLASQLVSRAVTGIDTILMTQRLPEPLGEGFF 1659
Db 1931 KSMVLLHLSQRY-----IFEPKNDRLSS--VTMPNVARQITLERS-----VGYY 1975
QY 1660 ANFVLPKYDPABHG-----DERWFKIHIGNVGNVGRQPYYS--GMLS-----DTSE 1704
Db 1976 RNI-----YQPPGNSAVIQDFTEDGHLHTFVLG--TGRRVYIKYKLSKLAETLYDTTK 2029
QY 1705 TSMTLFVPAEYGM-----HEGVLGVGVQKI-----TY 1734
Db 2030 VSFT-----YDGTAGLMTINLQNEGFTCTIRYQGLIDROI FRFTBEGMWNARFDNY 2085
QY 1735 DMTWESAPFYDTRKQOQVFLINDA-----DHDSGMTQO-----G 1768
Db 2086 DNS-----FVTSNQAV--INETPLDILYRDDVSGKTEQFGKFGVIYDINOIITA 2137
QY 1769 IVKNIKKYKGLNLSIATGYSAPMDFNSASALYYWELFYTPM--MCFQRLLOEQKOFDEAT 1827

Db 2138 VMTHKHFDAY-----GRMKEVOYEIERSLMTYMTVOYDNMGVRVKELKVGFIANTT 2190
QY 1828 QWINVVVPAGYIVNGEITAPWICRPLEETTSW-----NANPLDAIDPPDAVAQNDPMH 1881
Db 2191 RY-SVEYDADQOLQTVSI-----NDKPL-----WRYSYDLNGN--LHLLSPGNSARLTPLR 2238
QY 1882 YKIATFMRLLDQILRGDMAYRELTRDALNAKMWYVTRLELLGDEPDYDYSQQWAAAPSL 1941
Db 2239 YDIR-----DRITRLGDOVK-K-MDEDF-----LRQGGDIFEYNSAGILLIKAY 2281
QY 1942 SGMASTQVQAAQQODLTMLGRGYSKLNRTANSVLGLFLPEYNPA-----LTD 1989
Db 2282 NRAGWSVRYXIDG-----LGR-RVSSKSSHHLQFFYADLTNPKVTHLVNHSSEITS 2336
QY 1990 YWQTLRLRLFNLRN-----LSID--COPLSL-----AIYABPT---2021
Db 2337 LYDLOGLHFLAMELSSGDEFFVIACDNIGTFLAVSGTGLMIKQILYATAYGEIYMDTNPF 2396
QY 2022 -----DPKALLTSMVQASQ-----GGSAVLPGTSLYR 2049
Db 2397 QIIIGYHGLYDP--LTKLVHMGRRDYDVLAGRMTSPDHBLWHLSSSNVMPFLYMPK 2453
QY 2050 FPMVLERTRNLVAQLTOFGTSLLSWA-----EHDDADELTTLILLQCGMELATQSI 2099
Db 2454 NNPISNSQDIKCFMTDVNSWLLTFGFLHNVIPGYPKPDM-----AMEPSYELI 2504
QY 2100 RIQORTVEVDADIAVL-----ABSRSAQNRLKYQOLYDEIDINHGEQRAMSLDLAAAG 2154
Db 2505 HTQMKT-QEWDNSKSLGVCQVQKLFVTLERFDQLYGSTITSCQAPKTKKFASG 2563
QY 2155 QSLAGQVLSIAEGVADLVNPFVGLACGSRGWAALRASAVM-----SLSATASQY---2205
Db 2564 -SVFGKGVKFPALKDGRVTTDIISVANEDGRVAAILNHAHYLENLHFTIDGVDTHTYFKP 2622
QY 2206 --SADKISRSEAYRRRQWEIQRDNADGEVKMDAOLSLKIRREAAQM-----V 2255
Db 2623 GPSEGDIALGLSGGRRTLE-----NGNVTVSQINTVINGRTRRYTDIOQYGALCLNT 2677
QY 2256 EYQETQQAHTQAOLELQKFTNKA-----LYSWMRGK-----LSAIYYQFF 2297
Db 2678 RYGTTLDEKARVLELQARAVRQAWAREQQRLREGESGLRAWTEGBKQQLVSTGRVQGY 2737
QY 2298 DLTSQFCMLAQEARLRLTD--NGVTFIR 2324
Db 2738 D---GFFVISVEQY--PELSDSANNIHFMR 2762

RESULT 15

US-11-052-554A-171

; Sequence 171, Application US/11052554A

; Publication No. US2005028866A1

; GENERAL INFORMATION:

; APPLICANT: Sachdeva, et al.

; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; PROTEINS OF THERAPEUTIC POTENTIAL

; FILE REFERENCE: 30853/40359A

; CURRENT APPLICATION NUMBER: US/11/052,554A

; CURRENT FILING DATE: 2005-02-07

; PRIOR APPLICATION NUMBER: US 60/589,227

; PRIOR FILING DATE: 2004-07-20

; PRIOR APPLICATION NUMBER: IN 173/DEL/2004

; PRIOR FILING DATE: 2004-02-06

; NUMBER OF SEQ ID NOS: 763

; SOFTWARE: PatentIn version 3.3

; SEQ ID NO 171

; LENGTH: 2340

; TYPE: PRT

; ORGANISM: Rickettsia prowazekii

US-11-052-554A-171

Query Match 1.3%; Score 165; DB 7; Length 2340;

Best Local Similarity 18.1%; Pred. No. 0.052;

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:31:49 ; Search time 32.9071 Seconds
(without alignments)
4309.818 Million cell updates/sec

Title: US-10-754-115-45
Perfect score: 7901
Sequence: 1 MNSQDFSIETSLPKGGGA.....WFTVNEBNDTAAEVKKVKM 1474

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 100%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3569.5	45.2	1496	2 AH0447	insecticidal toxin
2	839	10.6	591	1 S09498	virulence-associat
3	835	10.6	591	1 S22664	virulence-associat
4	822.5	10.4	593	2 S15215	virulence-associat
5	709.5	9.0	563	2 S78224	virulence-associat
6	172.5	2.2	1404	2 E85509	hypothetical prote
7	172.5	2.2	1404	2 E90658	RhaS core protein
8	164.5	2.1	1399	2 A99720	RhaC core protein
9	161	2.0	1863	2 S46217	protein-tyrosine-p
10	160.5	2.0	1426	2 H64780	RhaD protein precu
11	158	2.0	3083	2 AH2493	hypothetical prote
12	156	2.0	2167	2 AF1489	cell wall-associat
13	155.5	2.0	1400	2 E90886	RhaS core protein
14	153	1.9	2036	2 T18995	hypothetical prote
15	151	1.9	1377	2 E86034	rhaA protein in rh
16	151	1.9	1394	2 H91236	RhaH core protein
17	151	1.9	1409	2 F91187	RhaH core protein
18	151	1.9	2314	2 T28698	hypothetical prote
19	149	1.9	1397	2 A85570	rhaC protein in rh
20	149	1.9	1398	2 B85549	hypothetical prote
21	149	1.9	1907	2 S08993	protein-tyrosine-p
22	148	1.9	1398	2 H90698	RhaD core protein
23	148	1.9	3972	2 S75251	hypothetical prote
24	147.5	1.9	682	2 C64898	rhaS protein - Rec
25	147.5	1.9	4199	2 S76412	hypothetical prote
26	145.5	1.8	632	2 F83387	copper resistance
27	144	1.8	709	2 S38241	hypothetical prote
28	144	1.8	1377	2 C65159	RhaA protein precu
29	144	1.8	1411	2 E65145	RhaB protein precu

30 142.5 1.8 1512 2 AH0439 probable membrane
31 141 1.8 1274 2 T37193 enamein matrix pr
32 141 1.8 1397 2 C64805 rhaC protein precu
33 141 1.8 2893 2 A64556 toxin-like outer m
34 139.5 1.8 1577 2 T30858 glucosyltransferas
35 139 1.8 828 2 G87584 hypothetical prote
36 139 1.8 2515 2 S47008 tenascin-like prote
37 138.5 1.8 1158 2 F90854 probable host spec
38 138 1.7 4427 2 P80637 polyketide synthas
39 137.5 1.7 656 2 G85731 Rhs element associ
40 137 1.7 2514 2 F81045 hemagglutinin/hemo
41 137 1.7 2628 2 T28651 hemagglutinin A -
42 136.5 1.7 1645 2 H85554 hypothetical prote
43 136.5 1.7 3566 1 A40701 tenascin-X precurs
44 136 1.7 775 2 C95921 hypothetical prote
45 136 1.7 1083 2 T23031 hypothetical prote

ALIGNMENTS

RESULT 1
AH0447
Insecticidal toxin complex [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AH0447
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0447
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1496 <KUR>
A:Cross-references: UNIPROT:Q8ZAV4; UNIPARC:UPI000000CDA67; GB:AL590842; PIDN:CAC93148.1.
C:Genetics:
A:Gene: YP03678

Query Match 45.2%; Score 3569.5; DB 2; Length 1496;
Best Local Similarity 49.3%; Pred. No. 3.8e-220;
Matches 748; Conservative 211; Mismatches 486; Indels 71; Gaps 29;

Qy 1 MNS-QDFSIETSLPKGGGAITGMEALTPGPDGMAALSLPLPISAGRGYAPFTLNY 59
Db 1 MENS-KQVAVAPLSLPKGGGATGNGDSLGPISGMAATLPLPISAGRVAPSLTLY 60
Qy 60 NSGAGNSPFLGWDGCVNMTIRRRTHFGVPHYDETDFTFLGPGGEVLVVDQPRD--ESTLQ 117
Db 61 SSGSGNGPFLGWLQGTMAIRRTNAQVPRYDEYDEFLAPNGEVNVAADPGSITERTEQ 120
Qy 118 GINLGATFTVGYSRLESFHSRLEYWQPKT-TGKTDFWLIYSPDQGVHLLGKSPQARS 176
Db 121 SLN-GEQSVIRYLPRIEFGNFRHIEYRPTNNSQAPFWLHSHSDGQKHCGLYSAAARIA 179
Qy 177 NPSQTOTPAOVLLEASVSSRGEQIYQVRAEDDTGCEADEI-----THHLOATAQRYLHI 231
Db 180 DPLHPEHTAEWLLSEVSLSGEHIGYQADEDEQIDEPSIYKAEKQNHPPASAOYLRK 239
Qy 232 VYGNRTASFTPLGLDGSAPSQADWLFYLVFDYDGRSNNLKTPTTAFSTTGWLCRODRFS 291
Db 240 VVYGNRQAAYELCYLT-QQPAPTSWLFSLIFDHGEYSNIAEQVPVILKGSWNFRQDAFS 298
Qy 292 RYEGFEIRTRRLCRQVLMYHHLQALDSKITEHNGPTLVSLRLILNYDSATSLVFRVR 351
Db 299 HFNYGFEVTRRLCQQLVLMYHNLKSGDEPDAQA-TLVSRRLYHQHDATVQLVGCCQ 357
Qy 352 VGHEDQGNVVTLPPELAYQDFSPRHHAHQPMVDMVLANFNAIQRMQLVDLKGEGPLGLY 411
Db 358 LAHEPDGTRSLPPLPEFDYQDFSTRDALGWQPLTDWAEFN--YQYQWVDLNGEGNPGMLY 415

A:Molecule type: DNA
A:Residues: 1-563 <NOR>
A:Cross-references: UNIPARC:UPI0000089A96; EMBL:X57096; NID:g46998; PIDN:CAA40380.1; PID:PID
R:Norel, F.; Pisano, M.R.; Nicoli, J.; Popoff, M.Y.
Res. Microbiol. 140, 455-457, 1989
A:Title: Nucleotide sequence of the plasmid-borne virulence gene mxfB from Salmonella ty
A:Reference number: A43996; MUID:90161559; PMID:2696057
A:Accession: S23713
A:Molecule type: DNA
A:Residues: 1-513, 'NITYLEKHGKGRILGDAVHPKGEAEMLF', 514-563 <NOW>
A:Cross-references: UNIPARC:UPI00001780F5; EMBL:X57096
C:Genetics:
A:Genome: plasmid
C:Superfamily: virulence-associated protein spvB
C:Keywords: DNA binding; transcription regulation

Query Match 9.08; Score 709.5; DB 2; Length 563;
Best Local Similarity 39.1%; Pred. No. 1.7e-37;
Matches 170; Conservative 60; Mismatches 138; Indels 67; Gaps 13;

Qy 9 ITLSLPGKGGAITGMCGEALTPTGDCGMAALSPLPISAGRGYAPFTLNNSGAGNSPF 68
Db 15 ITPPFLPKG-----GKALSQSGPDGLASITLPLISAERGFA----- 53

Qy 69 GLGWDCNV-----MTIRRTHF-----VPHYDETDTFLGPEGVLV-----VAD 108
Db 54 ---WRCCTAAVAAMALRRGLVLRDNEHCPPQPCVPOYNDSDFLGPDGEVLVQTILSTGD 110

Qy 109 QPRDESTLQ--GINLGATFTVTGVSRLSHFSRLSEYVQPKTKTKTDFWLLYSPDGVHL 166
Db 111 APNPVTSFAYGDSVFSPOSYTVTRYPQTESFRLSEYVQNSNG--DFWLLHDSNGILHL 169

Qy 167 LKGSPOARI NSPOTTQAOWLLEASVSSRGEQIYYOYRAEDDTGCEADRI THLQATAQ 226
Db 170 LKGTAAARLSDPQAASHTAQWLVEESTVPAGEHIYTSYLAENGDNVDLNGNEAGRD SAM 229

Qy 227 RYLHIVYVGNRTASSETLPLGLDGSAPOADWL FVLVFDYDGRSNNLKTPTPAFTTGWLCR 286
Db 230 RYLSKVQYGNATPAADL--YLWTSATPAVQWLFTLVFDYDGERGVDPQVPAPFTAQNSWLAR 288

Qy 287 QDRFSRYEYGEPIRTRLCRQVLMYHLQALDSKI TEHNGFTLVSRLLINLYDESATASTL 346
Db 289 QDPFSLNYGFEIRLRLCLCRQVLMFHH---FPDELGE--ADTLVSRLLLEYDENPILTQL 343

Qy 347 VFVRVGVGEODG-----NVVTLPPLELAYQDPSPRHHAHQWPMDLANFNFI-----Q 394
Db 344 CAARTLAYEGDYTRAPVNNMPPPPPPMGGNSRPSKKAIVEESKQIQALRYISAQ 403

Qy 395 RWQLVD--LRKEGLP 407
Db 404 GYSVINKYLRGDDYP 418

RESULT 6
E85509
hypothetical protein Z0268 [imported] - Escherichia coli (strain O157:H7, substrain EDL933
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85509
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85509
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1404 <STO>
A:Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00000D025B; GB:AE005174; NID:g12512977; H
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z0268
C:Superfamily: rhaF protein

Qy	1186	VOLT---DINDNQHLITLDALGRPIITRFWGTENGKMTGYSSEKASFSPPSDVNAIEL	1242
Db	862	PQLDRDYDNDNGOLIRISGPQESREYR--SDTGRLTGVHT-----TANLDIDI	910
Qy	1243	KKLPVACQCVAPESNMPVLISQKTENRLAEQDWQKLYNARIITEDGRICTLAYRRVQS	1302
Db	911	-----PYATDPAGNRLPD-----	923
Qy	1303	QKAIPQLISLNNNGPRLPPLSHLTITTDYDHPDPEQIRQQ---VVFSDGFGRLLOAARH	1359
Db	924	-----PELHPS--TLTA-----WPNRRTAEADAHVYRYDEYGRLEAKTDRI	963
Qy	1360	EAGHARORNEGSLIINVOHTENRWAVTGRTEYDNKCOPIRTQPYFLNDWRVYNSDSAR	1419
Db	964	PEGVIRHDE-----RTHHYHDSQRLVFFHTRIQ	993
Qy	1420	QEKAYADTHVYDPIGRKIVITAKG-WFRRTLTFTPMFTVNEDNDT	1465
Db	994	HGEQVESRYLYDPLGR-----TGKRVRRERDLTGWMSLSRKPET	1036

RESULT 7

E90658

RhSg core protein with extension [imported] - Escherichia coli (strain O157:H7, substrain C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

C;Accession: E90658

R;Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G. gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic reference strains

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: E90658

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1404 <RAY>

A;Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BA833660.1

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECe0237

C;Superfamily: rhsF protein

Query Match

Best Local Similarity 19.4%; Score 172.5; DB 2; Length 1404;

Matches 254; Conservative 137; Mismatches 427; Indels 489; Gaps 69;

Qy	327	PTLVSRLLINYDSAIASLTAVFVRRVCHGEQGNVVTLPPELAYQDFSPRHA-----	379
Db	51	PVLGAKVLPGETDLALGPPLFI-----LSRAYSYRTRTPAPGVGFGP	94
Qy	380	HWO--PMDVLANFNAIQRWQLVDLKGELPGLLYQDKGAWWYRSQRLGEIGSDAVTWKMM	438
Db	95	GWKAPFDI-----RLQTRD-----EGLIINDN-----GGRSIHFEPL	126
Qy	439	QPLSVIPSLQSNASLVNDINGDQDWWITFGFLRGYHSQRP-DGSWTRFTPLNALPVEY-	496
Db	127	FPGEI--SYSRSESF-----MLARG-GVAEQHSSQPLSALW-----QVLPEYVR	167
Qy	497	THPRAQLA--DLMG-----AGLSDLVLIGPKSVRLYANTRDGP-----AK	534
Db	168	LSPHMVLATNSLOQPMWILNWPVRPGEADVLPPEPPAYRVLTVGVVDGFGRTLAFHRAAE	227
Qy	535	GKDVQSGDITLVPYGADPRKLVAFSDVLGSGGAHLVEV-----SATKVT--CWPNLGRGR	588
Db	228	G-DV--AGAVTGVTDGAGRR-----FHLVL--TTQQAQRAEVRKQATSLSSPAGRSASS	279
Qy	589	FGQPIITLPGFSQPAPEF-----NPAQVYLADLDGSGPTDLIYVHTNRLDI	633
Db	280	LVPFDITLPA-----GTEYGADNGIRLEAVWLTHDPA--YPDELPAAPLARVYTTASGELRA	333
Qy	634	FLNKGSG---NGFAEPTVTLRPEGLRFDHTCOLQWADVQGLGVASLILSVPHMSPHH----	686
Db	334	YIDRSQTGVAGFA-----YDAEHAG-----RWVAHVYAGR	363

A:Accession: A99720
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1399 <HAY>
A:Cross-references: UNIPROT:Q8X9E8; UNIPARC:UPI00001633CB; GB:BA000007; PIDN:BAB34152.1;
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: ECs0729

Query Match 2.1%; Score 164.5; DB 2; Length 1399;
Best Local Similarity 19.6%; Pred. No. 0.067;
Matches 305; Conservative 166; Mismatches 570; Indels 513; Gaps 80;

QY 18 GGAITG-----MGEALTPTGPGDMAALSLPLPISACRGVAPAPFTLVNSGAGNSPFLGLG 71
DB 41 GGVTSGHVPNPLGAKVLPGETD--IALPGPLPILRTSYSYKT---KTPAPVGSGLPG 95
QY 72 WDCNV---MTIRRT-----HFGVPHYDETDTFLGPEGEVLVADQPRDESTL----- 116
DB 96 WKMPADIRLQLRDNLTILSDNGGRSLYFE-HLFPGEDG-----YSRSESLMVRGGVA 147
QY 117 ---QGINLGATFTVTGVRSLRLESFSLRYWQPKTKGTFWLIYSPDGQVHLIG-----KS 170
DB 148 KLDEGHRLAALWQALPEELRSLPH-----RYLATNSQGPWMLLWGCERY 192
QY 171 PQARISNPSQTTQTAQWLLASVSRRGE-QIYYQYRAEDDTGCEADRIITHLQATAQRYL 229
DB 193 PEADEVLPAFLPPYR--VLTLGLVDRFGRTQTFHREAAFEFGS-----EITGVDGAGRHR 246
QY 230 HIVYGNRTASE-----TLPGI-----DGSAPSQADWLFLYVFDYGER 267
DB 247 LVLTTQRAEAREARQAISGGTESAPFDTLPGYTEYGRDNGIRLSAVMLTH-----DPEY 302
QY 268 SNNLKTPP-----AFSTGSMUCRODRFSR-----VEYGEIRTRRLC-----RQVLMTHHL 314
DB 303 PENLPAAPLVRYGWTPRGELAVVYDRSGKQVRSFTYDDKYRGRVMAHRTGRPEIRYR- 361
QY 315 QALDSKITEHNGPTLVSLRLIYNDESALFLVFRV-VGHEQ--DGNVVTLPPELAYQD 372
DB 362 -DSGRVTEQLNPAGLS-YTYQYKDRITITDSLNRRVLHTQEGGLKRVVKE----- 414
QY 373 FSPRRHAWQPMVDLANFNFAIRQWLVLKGEGLPGLLYQDKGAWWYRSARQLGEIGSDA 432
DB 415 -----HADGSVTSQFQDAVGR-----LRAQ-----TDAAGRTTESPDV 448
QY 433 VTWEKQPLSVIPSLQSNASLVNDINGQLDWITGP-GL-----RGYHSQRPDG 481
DB 449 VT-----GLITRITTPDGRASAFYNNHHSOLT-SATGPDGLEIRREYDEWGLIOETAPDG 503
QY 482 SWTRFTPLNALPVEYTHPRAQL-----ADLMGAGLSDLVLGPKSVRLVYANTR-DGPAKGK 536
DB 504 DITRY-----RYDNPHSDLPQATEDATG---SRKMTWRSRYGQLLSFTDCSGYVTRY 552
QY 537 DVVQSGDITL-----VPFGADPR-KLVAFSDVLGSGQAHLEVEVSATKVTCPNLRG 587
DB 553 DHDREFGQWTAHREGLSQRAYDSRGOLIAVKDTQGHETRYEYNAAGDLTT----- 604
QY 588 RFGQPIITLPGSPQATFENPAQVYLADLDGSGPTDLIYVHTNRDLIDFLNKGW--GFAEP 645
DB 605 -----VIAPDGRNGTQYDAMGKAICTTQGGLTSMYDRAAGRVIRLTSENGSHTTFRYD 659
QY 646 VTLRFPEGLRFDHTCOLQADMDVQGLGVASILLSPHMSPHHWRCD-----LTN-----M 694
DB 660 VLDRLIQETGDTGRTQRYHDLTG-----KLIRSDDEGLVTHWHYDEADRLHRTVKGETA 715
QY 695 KPWLLNEMNNMNVVHHTLR-YRSSSQFWLDEKAAALTTGTPFCYLPFPPIHLLWQETED 753
DB 716 ERWQYDERGMLTDISHISEGRVTVHYGYDEKGR--LTGERQTVHHQTEALLWQETRH 773
QY 754 EISGNKL-----VTLIRVARGAWDG-----REREPFRGQYVEQ 786
DB 774 AYNAQGLANRIPCIPDSLPAVEWLTYSGWLAKMGLDTPLVDFTRDLRHLRTRFRGYEL 833

QY 787 TDSHQLAQGNAPERTPPALTKNWYATGLPIDNALSTEYWRDQAFAGFSFRFTTQDNK 846
DB 834 TTAY-----TPAGQLQSQHLSNL-----QYDRD-----YTNDN- 862
QY 847 DVPLTPEDDNSRVWFRNALKGQLLR-----SELYGLDDSS--TNKH-----VPYT 888
DB 863 -----GELIRSSPRQTRSYSDSGRLTGVTHTTAANLDIRPYA 902
QY 889 VTEFRSQVRRLQ-HTDSRYPLVASSVVESSNYHYERIASDPQCSQNTILSSDRFGQLKQ 947
DB 903 TDPAGNKLRLPDEILHDPSTLS--MMPDNRIARDAHY-----LYRYDRHGRLETEK 948
QY 948 LSVQYPRRQOPAINLYPDTLPDKLLANSYDQDQQLRLETTYQSSSWHHLTNNT----- 999
DB 949 -----TDLIPEGVIRT--DDERTHR--YHYDSQHLVHVTRTQVEEPL 987
QY 1000 --VRVGLPPOSTSDIFTYGAENVPPAGLNLLEL-----SDKNSLTADDPREYILG 1048
DB 988 VESRYLYDPLGRRVAKRVMRRERDLTGMMSLSRKPQVTWYWGDCGDRLLTIQNDTR- 1043
QY 1049 QOKTAYTDGQNTPLQPTTRQALIAFTETTVFNQSTLSAFNGSIPSKLSLTLEQAGVQ- 1107
DB 1044 -IQTIYQGSFTPLIRVETATGELAKTORR-----SLADALQOOSGED 1085
QY 1108 -----QTNLPFRTGED-KVMVA----- 1124
DB 1086 GGSVVPPVLVQMLDLRLESEILADRVSESRRLWASGLTVAQMSQMDPVYTPARKIHL 1145
QY 1125 ---HHG-----YTDYGTAAQFWRP-QKQSNLTOLTG-----KITLWD 1157
DB 1146 YHCDHRLPLALISKEGATWCWCAEYDEWGNLLAENPHQLQQLRPLGQQYDEESGLYNN 1205
QY 1158 AN-YCVVVQ---TRDAAGLTTSKDYDRF---LTPVQLTDINDNQHLITLDALGRPI 1208
DB 1206 RHRYDPLQGRYITQDDPIGL---KGNWNYQYPLNPVQY-----IDSMG--LA 1248
QY 1209 LRFWGTENGWMTGYSS--PEKASFPSPSDV-NAAIELKKPLPVAQCQVYAPESMMPVLQS 1266
DB 1249 SKYGHLLNG--GYGARENKPPTPDPSKFPDIAKQLRLPYIDQA-----SSAPNVFKT 1299
QY 1267 TFRNLABQWKLQYNARIITEDGRICITLAYRRVQSQAKAIPOLISLNLNGRPLP 1320
DB 1300 FFRALSPDY-----TLYCRKWKV-----PNLTCTCPQDDPQYP 1332

RESULT 9

S46217
protein-tyrosine-phosphatase (EC 3.1.3.48) type sigma precursor - rat
N;Alternate names: leukocyte common antigen-related phosphatase
C;Species: Rattus norvegicus (Norway rat)
C;Date: 07-May-1995 #sequence revision 03-Nov-1995 #text_change 09-Jul-2004
C;Accession: S46217; S51174; A49104
R;Zhang, W.R.; Hashimoto, N.; Ahmad, F.; Ding, W.; Goldstein, B.J.
Biochem. J. 302, 39-47, 1994
A;Title: Molecular cloning and expression of a unique receptor-like protein-tyrosine-phosphatase
A;Reference number: S46216; MUID:94347119; PMID:8068021
A;Accession: S46217
A;Status: nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-1863 <ZHA>
A;Cross-references: UNIPROT:Q64605; UNIPARC:UPI0000177058; EMBL:L11587
R;Goldstein, B.J.
submitted to the EMBL Data Library, February 1993
A;Reference number: S51174
A;Accession: S51174
A;Molecule type: mRNA
A;Residues: 1-1788, 'G', 1790-1863 <GOL>
A;Cross-references: UNIPARC:UPI000008AD4; EMBL:L11587; NID:G205134; PIDN:AAC37656.1; PI
R;Yan, H.; Grossman, A.; Wang, H.; D'Eustachio, P.; Mossie, K.; Musacchio, J.M.; Silvenn
J. Biol. Chem. 268, 24880-24886, 1993
A;Title: A novel receptor tyrosine phosphatase-sigma that is highly expressed in the ner
A;Reference number: A49104; MUID:94043351; PMID:8227050
A;Accession: A49104

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-596,'R',598-603,'I',967-1788,'G',1790-1863 <YAN>

A:Cross-references: UNIPARC:UPI000005ED5

A:Experimental source: brain

A>Note: sequence extracted from NCBI backbone (NCBIP:139669)

C:Superfamily: leukocyte antigen-related protein; fibronectin type III repeat homology;

cyg

C:Keywords: alternative splicing; glycoprotein; phosphoprotein; phosphoric monoester hyd

F:1-26/Domain: signal sequence #status predicted <SIG>

F:27-1863/Product: protein-tyrosine-phosphatase #status predicted <WAT>

F:149-209/Domain: immunoglobulin homology <IMM1>

F:248-300/Domain: immunoglobulin homology <IMM2>

F:318-400/Domain: fibronectin type III repeat homology <FN3A>

F:413-499/Domain: fibronectin type III repeat homology <FN3B>

F:511-592/Domain: fibronectin type III repeat homology <FN3C>

F:1244-1863/Domain: leukocyte common antigen cytosolic domain homology <LAC>

F:1331-1552/Domain: protein-tyrosine-phosphatase homology <PTP1>

F:1504/Active site: Cys (phosphotyrosine intermediate) #status predicted

F:1510/Binding site: substrate phosphate (Arg) #status predicted

F:1795/Active site: Cys (phosphotyrosine intermediate) #status predicted

F:1801/Binding site: substrate phosphate (Arg) #status predicted

Query Match 2.0%; Score 161; DB 2; Length 1863;

Best Local Similarity 19.1%; Pred. No. 0.18;

Matches 336; Conservative 188; Mismatches 613; Indels 620; Gaps 87;

Qy	9	ITELSLPKGGG---	AITGMGALTPGDPGMAALSPLPISAGRGYAPAFILNYSNGAGN	65
Db	313	ITVSLPKAPGTPVVTENTATSIITVDSGN-----	PDVVS-----YYVIEYKSKSQD	360
Qy	66	SPFGLGDCNMTIRRRTHFGVPHYDETDTL-----	GPEGVLVY-----ADQP	110
Db	361	GPYQIKED---ITTRYSIGGLSPNSEIVEIWVSANVSIQGPSSSVTRTGEQAPASAP	417	
Qy	111	RDESTLQINIGAFITVGYRSLRSHFSRLYEQPKTKGTDWFLIYSPGQVHLGKS	170	
Db	418	RN---VQARMLSATTMIVQWEEPEPN-----	GLIRGYRVYVYTWEPHEPVGNN	462
Qy	171	POARISNSQTTQTAQWLEASVSRRGQIYVQVRAEDDTGCEADEITHLQATAQRYL--	229	
Db	463	QKHNVDLSLIT-----	VGSLLEDEYTVRVLAFTSVGDGPLSDPIQVKTQGVV	512
Qy	230	-HIVYGNRTASLTPGLDGSAPSQADWLFY-LVFYDYGERSNNLKTTPAFSTTGSWLCRQ	287	
Db	513	GQPNMLRAEAKSETSIGLSWAPRQESVIKVELLPREGDRGRCR--TFDPTAFVV-E	569	
Qy	288	DRFSRYEGPEIRTR-----	LCQVLMYHHLQA-----LDSKITEHNGPTLVS	331
Db	570	DLKENTYAFRLAARSPOGLGAFVAVQCRT-----	LOAKFSAPPQDVKCTSLRSTAI--	622
Qy	332	RLIL-----	NYDESAIA-----STLVFVRRVGHEDGNVVTLPPLLEYQDPS	374
Db	623	-LILLEALEKTEYKTVAVATEVCGPSSPVVVR-----	TDEVPSAPPKRYEAEALN	676
Qy	375	PRH-HAHQPMVDLVANFNALQRWLQVLKGBGLPLGLLYQDKGAMWYRQAQRLGEIG-SDA	432	
Db	677	ATAIRVLWRSPTPGHQHQRGQVGVHYVVRMEGTEA-----	RGPRIKIDIMLADA	725
Qy	433	VTEKMQPLSVIPSLQS-----	NASLVVDINGDQ---QLDWVITGPGLRGYS-----QRP	479
Db	726	---QEM-----	VITNLQPTAYSITVAAYTMKGDGARSKPVVVTVTKGAVLGRPTLSVQQT	778
Qy	480	DGS-----	WTRFTPLNALPVEYTHPRAQLADLMGAGLSDLV-----IGPKSVRLY	525
Db	779	EGSLIARW-----	EPADAAE-----DPVLGYRLQFGREDAAPATLELA	817
Qy	526	ANTRDGFA---KGKDW-----	QSGDITLPVFGADPRKLVAFSDVLGSGQAHL	570
Db	818	AWERRFAAPAHKATYVFLAARGRAGLGEESAALSIPEDAPR---GFPQILGPAG---	871	
Qy	571	VEVSATKVC-W-----	PNLGRG---RFGQPTILPGFSQPATE-----FNPAQVYLADLD	616

Db	872	NVSAGSVILRWLPPVPAEGNGAIKIYTVSVREAGTGPATETELAAAAQCAETALTIQ	930	
Qy	617	GSGETDL-----IYVHTNRDLIFLKNKNG-FAEPVTLR-----	649	
Db	931	GLRPETAVELRVRAHTR-----	GGPFSPLRLRYLARDPVSPKFKVIMKTSVL	982
Qy	650	-----FPEGLRDHTCQLQ-----	MADVQGLGVASLILSVPHMSPHHWRCDLTNNKMWLLNE	701
Db	983	LSWEPFDNYSPTPKYQIYNGLTLDVGRTRTKKLI---THLKP-----	TFYNFVLNTR	1033
Qy	702	MNNMGVHTLURYRSSQFWLDEKAAA-----	LTGTOTPV-----CYLP--	740
Db	1034	GSSLGGLQOQTVATATFANMLSGKSPVAPKPDNDGSIIVVYLDGQSPVTVQYFIVVPLR	1093	
Qy	741	-----PPIHTLQWTEDEISGNKLV-----	TTLYRARGA-----	770
Db	1094	KSRGQGFPI-----	LLGSPEDMDLEELIQLDLSRLQRRSLRHSRQLQEVPRPYIAARFSILPA	1149
Qy	771	--WGREREPFRGFGVSGQTDSHOLAQGN-----	APERTPPALTKNWYATGL-----	814
Db	1150	VFHFGNQKQYGGF-----	DNRGLEPGHRYVFLVAVLQKNEPTTFAASFFSDFDOLDNPD	1203
Qy	815	--PVIDN-----	ALSTEYRDDQAFAGFSRFTTWQDNKDV--P	849
Db	1204	POPIVDGEGLIWIGPVIAVFIICIVIALLLYKNKPDSSKKDSEPTKCLNNADLAP	1263	
Qy	850	LTPEDDNRVWFNRALKQQLRSBLYGLDDSTNKHVYTVTFEFSQVRRLQHTDSRYVPL	909	
Db	1264	HHPKDPVEMRRINFQTPGML-----	SHPPITIDMAEHMERLKANDSL----	1306
Qy	910	WSSVESRNYHYRIASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQOQAINLYPDTL--	967	
Db	1307	-----KLQSEYESI--DP-----	GOOFTWEHSNLEANKPKRYANVJAYDHSRV	1349
Qy	968	-----PKLLANSYDDQROLRLTYQQSSWHLLTNNTVRVLGLPDSRSDIFTYGA	1018	
Db	1350	LQPLEGMSYINAVYDGVRRQ-----	NAYIATQGLPEFGDFWRVWV	1395
Qy	1019	ENVPAGGLNLELLSDKNLSIADDRPREYLGOQKTAITDQNTTLPQTTRQALIAFTET	1078	
Db	1396	EORSATVMMTRLEKSRVKDQ-----	YWNRG-----TETY	1428
Qy	1079	VFNQSTLSAFNGSIPSDKLSLTLEAQYQOQNYLFPRTG-EDKVVVAH-----	HGVT	1129
Db	1429	GFIQVT-----	LLDTMELATFCVTRFSLHNGSGSEKREVRHFQFTAWPDHGV	1476
Qy	1130	DYGTA-AQFWPQKQSNLTOLTKITLIWDANY-----	CVVV-----QTRDAAGLT	1173
Db	1477	EYTPFLAFLRRVTCNPPDAGPVVHCSAGVGTGCFIVIDAMLERIRTEKTVDVYGHV	1536	
Qy	1174	TSKYDRWRFLTPVQLTDINDNHLITLDALGRPITLRFWGTENGKMTGYSSEKASFSPP	1233	
Db	1537	TLMSQRNYM--VQTEQDQYSFHEALLEAVG-----	CGNTEVPARSILTYTIQKLAQVEP	1588
Qy	1234	SDVNAALTELKPLPVAQCOVYAPESWVPLSQKTFNRLAEQDWKLYNARIIT-----	1286	
Db	1589	GEHVTGMELE-----	PKRLAS---SKAHTSRFTASLPCNK	1621
Qy	1287	-----EDGRICTLAVRRWVQSKAIPQLISLNLNGPRLPHSLTLTTDHYDHP	1335	
Db	1622	PKNRLNVLPESSRVC-LQPIRGVSGDYI-----	1651	
Qy	1336	EQIIRQOQVVFSDGFRLLQAAARHEAGMARQNRNEDGSLIINQVHTENRWAIVTGRTEYDNK	1395	
Db	1652	-----NASFIDGY-----	RQKAYIATQ---GPL---AETTEDFW---RALWENN	1687
Qy	1396	QPIRTYQPIYFLNDRVYVNSDARQ-----	EKEAYADTHYDPIG-----REIKVIT	1442
Db	1688	STIV-----	VMLTKLREMGREKCHQYVPAERSARYQYFVVDPMFMAEYNNPQVILREFKVT	1742
Qy	1443	AKGWFRTL-----	FTPW	1455
Db	1743	ARDGOSRTVROFQFTDW	1759	

QY 1421 EKEAYADTH-VYDPI-GREI 1438
| : ||| : |||
Db 1210 ESGLYNNRHYDPLQGRYI 1229

RESULT 11
AH2493
hypothetical protein all71128 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120a1
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AH2493
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, S.; Tanaka, K.; Shimizu, T.; Takazawa, M.; Yamada, M.; Tabata, S.; Igarashi, S.
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A;Reference number: AB1807; MUID:11595285; PMID:11759840
A;Accession: AH2493
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-3083 <KUR>
A;Cross-references: UNIPROT:Q8YL10; UNIPARC:UPI00000CEEDB; GB:BA000020; PIDN:BA078212.1;
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all7128
A;Genome: plasmid

Query Match 2.0%; Score 158; DB 2; Length 3083;
Best Local Similarity 18.8%; Pred. NO. 0.66;
Matches 276; Conservative 163; Mismatches 507; Indels 520; Gaps 73

Qy	1	MQNSQDPSITELSPKGGGALTGHEAL-TTGPDDG-----MAALSLPPIASAGRYA	52
Db	725	IEAAEYVNAASAGAGGGIITGLNLKODTPGDGKVRGNEFVQLLNNPIEMFADSLV	784
Qy	53	PAFTLNTNSGAGN-----SPFG-----LWDCNVMTIR-----	80
Db	785	QAYIMAYAKVAGKVVKRIESPKVLLGPYKVSETPPQHLHLATDGGGNLRNLNMPNAAA	844
Qy	81	-----RRTHFGV-----PHYDET	94
Db	845	REIINTEDGAEVFTVTDGKLTVSAFNIPQTVSGVSKIIADGGTKNDTIBKPDIEISA	904
Qy	95	TFIIGPEGEVLVADQPRDESTLOGINLGAITVTCYRSRLSHSRLEWYQPKTKGTD	154
Db	905	DLKGGAGEDLYGSGSD--TIRG--GAD-----WDRLYGDDR	940
Qy	155	WLIYSPDQVHLLGKSPARISNPSQTQTQAQWLLLEASVSRGEOYYQYRAEDDTGCEA	214
Db	941	F-VYGDGDDWLDG-GAGADILGGAGFDT-----ASYTSATSAINLVTQVSTGDAA	992
Qy	215	DEITHHL-QATAQRYLHIV-----YYGNRTASETLPG-----LDG---	248
Db	993	DDVFQSEIQIVGSRYYDITLIGDENNEFDGGEKDNFISGGAGDDRSLSPGWGDVDTGGTG	1055
Qy	249	-----SAPSOA-----DWLFYLVDPDGERSNNLKTTPAFSTTGSWLCRQD	288
Db	1053	TDTLVIVDYSLSLFTQAVAWSELDPNTSDMFVVANAIGTGA-PIKTD--INVS	1103
Qy	289	RFSRYEYCFEIRRLCRQVLMYHHLQALDSKITEHNGFTLVSRILNLYDSEATSLTVF	348
Db	1104	-----YHATLSAD-----GLTVAGSGILGSGNS--GNQGLW	1133
Qy	349	VRRVGHQDQGNVTLPPLLEYQDPSPRHHAH--WQPMDV--LANFNALQRWOLVDLK--	402
Db	1133	VKKI-HSSDPAVRVPIPNQV-YQPLLEDGSKVWMSQDSTWIANTNGTQVRQLTKLSIN	1199
Qy	403	-----GGLPLGCLLYQDKAWVRSAORLGEI-----GSDAVTWEKQWPLSVIPSLQSN	450
Db	1191	IGYGDGYLATISDGGSTIAWLRKRNNDKTYTIFIANADGKNLRQIN-----IFTSGG	1244
Qy	451	ASLVYDINGDG-QLDWVITGPGLRGYHSRPPGSGWTRFTPLNALPVEXTHPRAQLADLMCA	509

RESULT 12

AF1489

cell wall-associated protein

C:Species: *Listeria innocua*

C;specrev: L13CER11100Cua
C;Date: 27-Nov-2001 #sequence revision 27-Nov-2001 #text change 09-Jul-2004

C; Date: 27-NOV-2001 #sequence
C; Association: AF1490

C;Accession: AFI489

R;Glaser, P.; Frangeul, L.

.; Dominguez-Bernal, G.; D

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A: Authors: Krest, J.: Kubn

A; AUCIOB: KLEIC, U.; KALL
OK C : schlüßter T : sim

ok, C.; Schueter, T.; Sim

A;Title: Comparative genom

Db 276 LSDSAPPDTLPG-----TEYGPDRGIRLSAVWLTHDPAYPESLPAGPLARYTYTBAGELL 330
QY 633 IFINKS-----GNGFAEPVTLRPFEGRLRPDHTCOLQADVQGLGVAS 674
Db 331 AVYDSNTQVRAFTYDAQHPGRWAHYRAGRPEMY-----RYDDTGRV-----VEQLNPAG 382
QY 675 LILSVPHMSPPHWRCDLTNMKPLLNNMNNMGVHHTLR--YRSSSQFWLDEKAAALTTG 732
Db 383 LSRYQYEQDRITVTDSLNRREVLTATE--GGAGLRKRVVKELADGVSVTHSGYDAAGRLTA 440
QY 733 QTPVCVCLPPIHTLQTEDEISGNKLVTLRYARG-----AWDGREREB----- 778
Db 441 QT-----DAAGRTYGLNVVSGDITDITPDGRETKEFYNDGNOL 481
QY 779 -----RGFG-----YVEQT-----DSHQLAQGNAPERTPPALTKN-----W 809
Db 482 TAVVSPDGLSRRAYDEPGRVLSVSRGCDVIRYAYDN- PHSELPAITTDATGSTRTQMTW 540
QY 810 --YATGLPVID--NALST--EYWRDQAFA-----GFSRPTTWQDNKDVLPED-----DN 856
Db 541 SRYGQLLAFTDCGYQTRYEDRFQGMATAVHREEGIS--RYRY--DNRGRLTSVKDAQGHE 598
QY 857 SRYWPNRA-----LKGQLLSSELYGL-----DGS 880
Db 599 TRYENAGADLTAVITPDGNRSETQYDANGKAVSTTQGLTKSMYDLAGRITTLTNENG 658
QY 881 TNKHVPYVTEPRSQVRRLQHTDSRYPLVMSVSVSRNYHYRIASDPQCSQN--ITL-- 936
Db 659 SRSEFTYDALDELVOORGF--DGR-----TQRHYDLTKLQSEDEGLVTLWH 705
QY 937 --SSDRP-----GQPLKQLSVQYPRR-----QQPAINLYPDTLPDKLLANSY 976
Db 706 YDESRLTHRTWNGSPAEQW--QYDEHGWLTEISHLSEGHQVAVHY-----GY 751
QY 977 DDOORQL--RLTYQSS-----WHLTNNYTVRLGL-----PDS--TRSDIPTYGAEVPA 1023
Db 752 DDKGLAGERQTVHNPETGELLQWQHETBAYNEOGLANRVTPDSLPVREWITYGS----- 806
QY 1024 GGLNELLSKNSLIADDPREYLQOKTAYTDGQNTTPTQPTTQALIAFTETTTFVFNQS 1083
Db 807 -----CYLAGMKL-----GCTPLVEFTDRL-----HRE 830
QY 1084 TLSAFNGSIPSKLSTLEQAGYQNTYLFPRTGDKVVAHGHVTDYGTAAQFWRPK- 1142
Db 831 TVRSPGNN--AYELTSTVTPAGHLQSORL-----NSQVYDRYDNDNGDLVRISGPRQT 883
QY 1143 --QSNTQLTKI--TLWDANYCVVQTRDAAG-----LTTSA 1176
Db 884 WEYGSATGRLESVETLASDLDLIRIPYAT--DPAGNRLPDPPLHDPDSTLTAMPDNRIABDA 942
QY 1177 KYDWRFLTPVQLTDIN-----DNQHLITL----- 1200
Db 943 HVVYRHDEYGLTEKTDRIPAGVIRTDDBRTHHHYDSQHLRVFYTRIQHGEPLVESRYL 1002
QY 1201 -DALGRPITLRFWTEGKMTGYSS-----PEKASFPSPSDVNAALTELKPLPVAQCQ-VY 1254
Db 1003 YDPLGRMAKRVRRER--DLTGWMSLSRKPEVTWYGMGDRLLTTVQ-----TDTTRIQTVY 1057
QY 1255 APESHWPVLSQKTFNRLAEQDWKLYNARIITEDEGRICTLAYRRWVQSKAIPQLISLN 1314
Db 1058 EFGSFTPLIRVETENGEREKA--QRSLAETLQOEGS-----E 1093
QY 1315 NGPRLP--PHSLTTLTDYDHDPEQIQVQVFSDFGRLQLQAAARHEAGMARQRNED--G 1371
Db 1094 NGHGVVFPFAELVRLDLR-----EEIRADRVSSESRAWLAQCGLTVEQ--LARQVEPEYTP 1148
QY 1372 SLIINVQHTENR-----WAVTGRTEYDNKGPRTVTPYPLND--WRYVEND 1416
Db 1149 ARKVHFYHCDHRGLPLALISEDNTAW-----RGEYDEWGNQNLNENPYLHPVRLPQQ 1204
QY 1417 SARQKEAYADTHVVDPI--GREI-----KVITAKGW 1446
Db 1205 HDEESGLYNNRNYDPLQGRYITQDPIGLAGGW 1238

RESULT 14

T18995

hypothetical protein C06B8.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004

C:Accession: T18995

R:Steward, C.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19057

A:Accession: T18995

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-3036 <WIL>

A:Cross-references: UNIPROT:O17575; UNIPARC:UPI000017B740; EMBL:Z81463; PIDN:CAB03852.1.

A:Experimental source: clone C06B8

C:Genetics:

A:Gene: CESP:C06B8.7

A:Map position: 5

A:Introns: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1325/3

Query Match

1.9%; Score 153; DB 2; Length 3036;

Best Local Similarity 18.7%; Pred. No. 1.3; Mismatches 459; Indels 428; Gaps 65;

Matches 241; Conservative 164;

QY 335 LNYDESATIASTLVFVRRYVGHEQDGNVVTLPPELAYQDFSPRHHAAHQWMDVLANFNATQ 394

Db 118 INYDSEMPENFRLV-----DGPVTVQGRLOVQFRD-----RWRSVCTM-----VT 156

QY 395 RWQLVDLKGELGPLYQDKGAW-WYR-----AVTWEKQPL--SVIPSLQSNASLVDIN--GDGQ 461

Db 157 NWTSID-TGTACRSNGYSDGGFWKFRNRNDTYPFVMPKPDCHGAANKLWDCPAFSNPQK 215

QY 429 -----GSD-----AVTWEKQPL--SVIPSLQSNASLVDIN--GDGQ 461

Db 216 IRLSENLCQSDDDIGYCWGPPPTFGWARHWKQILNSPHYVNSDPLVAVNRESR 275

QY 462 LDWV-ITGPGILRGYHSQRPDGSWTRFTP--LNALPVEYT-HPRAQLADLMGAGLSDLVLI 517

Db 276 LEFVDILVAGYDVGNKNTTSALYIEGVPPINWGLRIEHSARDGLQLLDANGPA----- 328

QY 518 GPKSVRLYANTRDGFAGKQVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLEVSATK 577

Db 329 -----ITANSTFSYNRGHGI-----SVVNTTDARIF-INNTK 359

QY 578 VTCWPNLGRGFGQPIITLPGFSQPATHENPAQVVLADLD-----GSGPTDILYVHTNR 630

Db 360 I-----QGNWGDGI-----WYKOQTGVN-----LIDYGMERRRSIGSRLE---EQKPR 400

QY 631 LDI FLKSGNGFABPVTILRPFEGRLRFDHTCOLQADVQGLGVAS-----LILSVPHMSPH 685

Db 401 IDMCAEH-----RVDDNHFFPHLIAVNLKRTYLDLAQPAICWMTVSLPRLPY 449

QY 686 HWRCDLTNMKPLLNNMNNMGVHHTLYRSSSQFWLDEKAAALTTGTPVCYLFPFPHHT 745

Db 450 -----TVSIQWLHIRDNPQTARTTLIACDSNN--VDENSCSTPRFPIRNEIFP--- 498

QY 746 LMQTEDEIISGNKLVTLRYARGAWDGREREPFGVVEQTDHQLAQGNAPERTPPAL 805

Db 499 ---QSILKSGKPLYLALEH---VLDGQA-----GYV-QGDVHLLFNIIHA-----SVL 541

QY 806 TKNWYA-----TGLEVIDN-----ALSTEYWRDDQAFAGFSFR----- 838

Db 542 DKAYYGLNVNTNCIIKTKTGNGVFANDIRERTALTNTVDLNDQGYAGFLVKGDAADWLNE 601

QY 839 ---FTTWQD-----NKDVPLTP-----EDDN 856

Db 602 TRIILNNMGDMNISYAGGSIMVNGTRIEKNRWRGAIIHYNQTLPLPLMYNEVIFKGRPSN 661

QY 857 SRYWPNRAK---GQLLRSELYGLDDSTNKGHPVYVTEPFSQVRRLOHTDSRYPLV--- 909

Db 662 NKFLPTIISENWGGLLVGNFCAYSNESNHSKTRIP-----INSVYPKLVPK 710

Search completed: February 16, 2006, 21:44:59
Job time : 41.9071 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:23:04 ; Search time 197.739 Seconds
(without alignments)
5259.201 Million cell updates/sec

Title: US-10-754-115-45
Perfect score: 7901
Sequence: 1 MQNSQDFSITSLSPKGGGA.....WFTVNEEDNTAAEVKKVKM 1474

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7901	100.0	1474	Q8GF99 PHOLU	Q8gf99 photorhabdu
2	7285.5	92.2	1475	Q7N7Y2 PHOLL	Q7n7y2 photorhabdu
3	6065.5	76.8	1476	Q7N7Z0 PHOLL	Q7n7z0 photorhabdu
4	6031.5	76.3	1476	Q93EP6 PHOLU	Q93ep6 photorhabdu
5	4627.5	58.6	1485	Q7N936 PHOLL	Q7n936 photorhabdu
6	4595.5	58.2	1485	Q85153 PHOLU	Q85153 photorhabdu
7	3984.5	50.4	1433	Q6XP56_YERFR	Q6xp56 yersinia fr
8	3971	50.3	1428	Q9F922_YENTR	Q9f922 serratia en
9	3638	46.0	1489	Q6QD24_YERPS	Q6qd24 yersinia ps
10	3591	45.4	1481	Q693A4_YEREN	Q693a4 yersinia en
11	3569.5	45.2	1496	Q8ZAV4_YERPE	Q8zav4 yersinia pe
12	3569.5	45.2	1516	Q8D1P6_YERPE	Q8d1p6 yersinia pe
13	3350.5	42.4	1401	Q93RN8_XENNE	Q93rn8 xenorhabdus
14	2277.5	28.8	1447	Q4ZP56_PSESM	Q4zp56 pseudomonas
15	2274	28.8	1446	Q87X46_PSESM	Q87x46 pseudomonas
16	1076	13.6	2439	Q4HWU2_GIBZE	Q4hwu2 gibberella
17	837	10.6	591	VRP2_SALCH	P17450 salmonella
18	835	10.6	591	V17486_SALTY	P74846 salmonella
19	832	10.5	591	V17486_SALTY	P74846 salmonella
20	829	10.5	591	V17486_SALTY	P74846 salmonella
21	826.5	10.5	593	V17486_SALTY	P74846 salmonella
22	826.5	10.5	593	V17486_SALTY	P74846 salmonella
23	709.5	9.0	563	Q4L1X7_9ENTR	Q4l1x7 salmonella
24	676	8.6	283	Q8KSN9_XENNE	Q8ksn9 xenorhabdus
25	651.5	8.2	2031	Q63XE6_BURPS	Q63xe6 burkholderi
26	502	6.4	1806	Q62N48_BURMA	Q62n48 burkholderi
27	457.5	5.8	166	Q6WCC1_9ENTR	Q6wcc1 serratia pr
28	386	4.9	2217	Q8TP72_METAC	Q8tp72 methanosarc
29	330.5	4.2	2379	Q72U39_LEPIC	Q72u39 leptospira
30	326.5	4.1	221	Q58FC3_SALIC	Q58fc3 salmonella
31	292	3.7	2306	Q8F107_LEPIN	Q8f107 leptospira

32 289.5 3.7 2554 2 Q72QR5_LBPIC Q72qr5 leptospira
33 279.5 3.5 2321 2 Q8F5B9_LEPIN Q8f5b9 leptospira
34 241.5 3.1 1826 2 Q987Z7_RHILO Q987z7 rhizobium 1
35 202 2.6 2370 2 Q82RE3_STRAW Q82re3 streptomyce
36 199 2.5 2060 2 Q6QBO0_STALU Q6qbo0 staphylococ
37 181.5 2.3 2364 2 Q82R58_STRAW Q82r58 streptomyce
38 180 2.3 1976 2 Q8D4R5_VIBU Q8d4r5 vibrio vuln
39 177 2.2 795 2 Q7NFS7_GLOVI Q7nfs7 gloeobact
40 174.5 2.2 1881 2 Q8TJS7_METAC Q8tjs7 methanosarc
41 172.5 2.2 1404 2 Q7AH10_ECO57 Q7ah10 escherichia
42 172.5 2.2 1404 2 Q8XED9_ECO57 Q8xed9 escherichia
43 171.5 2.2 2566 2 Q8TSE7_METAC Q8tse7 methanosarc
44 170 2.2 593 2 Q8GP37_PSEAE Q8gp37 pseudomonas
45 170 2.2 2082 2 Q9ACP4_STRCO Q9acp4 streptomyce

ALIGNMENTS

RESULT 1
Q8GF99 PHOLU PRELIMINARY; PRT; 1474 AA.
AC Q8GF99
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE TcdB2.
GN Name=tcdB2;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Ferry R.D.,
RA french-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family."
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RT "Genomic islands in Photorhabdus."
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AA017202.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR003284; Sal_SpVb.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
SQ SEQUENCE 1474 AA; 166338 MW; 9118D4E4914683DD CRC64;

Query Match 100.0%; Score 7901; DB 2; Length 1474;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1474; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQNSQDFSITSLSPKGGGAITGMGEALTPTGPDGMAALSPLPISAGRGVAPFTLNN 60
DB 1 MQNSQDFSITSLSPKGGGAITGMGEALTPTGPDGMAALSPLPISAGRGVAPFTLNN 60
QY 61 SGAGNSPFLGWCNCNMTIRRTTHFGVPHYDETDFTFLGPEGEVLVADQPRDESTLQIN 120
DB 61 SGAGNSPFLGWCNCNMTIRRTTHFGVPHYDETDFTFLGPEGEVLVADQPRDESTLQIN 120
QY 121 LGATFTVTGYSRLESHPFSRLYEQPTTKTDFWLIYSPDQVHLLGKSPQARISNPSQ 180
DB 121 LGATFTVTGYSRLESHPFSRLYEQPTTKTDFWLIYSPDQVHLLGKSPQARISNPSQ 180

```
QY 181 TTOTAQMLLEASVSRGEQIYYOYRAEDDTGCEADEITHHLOAQAORVLIHYVYGNRTAS 240
DB 181 TTOTAQMLLEASVSRGEQIYYOYRAEDDTGCEADEITHHLOAQAORVLIHYVYGNRTAS 240
QY 241 ETLPLDGSAPSQADWLFLYVDFDGERSNLKTTPAFSTTGSWLCRODRFSRYEYGFPIR 300
DB 241 ETLPLDGSAPSQADWLFLYVDFDGERSNLKTTPAFSTTGSWLCRODRFSRYEYGFPIR 300
QY 301 TRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLINLYVDESIAISTLVFVRVREGHQDGNV 360
DB 301 TRRLCRQVLMYHHLQALDSKITEHNGPTLVSRLLINLYVDESIAISTLVFVRVREGHQDGNV 360
QY 361 VTLPPELAYODFSRHHAWQPMVLANFNAIORMQLVDLKGSLGLYQDGGAMWYR 420
DB 361 VTLPPELAYODFSRHHAWQPMVLANFNAIORMQLVDLKGSLGLYQDGGAMWYR 420
QY 421 SAQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVLDINGDGLDWITGPGLRGYHSORPD 480
DB 421 SAQRLGEIGSDAVTWKMOPLSVIPSLQSNASLVLDINGDGLDWITGPGLRGYHSORPD 480
QY 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQDVQ 540
DB 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKQDVQ 540
QY 541 SGGITLTPVPGADPRKLVAFSDVLGSGQAHLYEVSATKVTCPNPLGRGRFGQPIITLPGFSQ 600
DB 541 SGGITLTPVPGADPRKLVAFSDVLGSGQAHLYEVSATKVTCPNPLGRGRFGQPIITLPGFSQ 600
QY 601 PATEFNPQAQVYLAIDLKSGGPTDLIVHTNRDLIDFLNKGNGFAEPVTLRFPGLRFDHTC 660
DB 601 PATEFNPQAQVYLAIDLKSGGPTDLIVHTNRDLIDFLNKGNGFAEPVTLRFPGLRFDHTC 660
QY 661 QLQMAVQGLGVASLILSVPHMSPHHRCDDITNKPMLNLNEMNNMNVHHTLRYRSSQF 720
DB 661 QLQMAVQGLGVASLILSVPHMSPHHRCDDITNKPMLNLNEMNNMNVHHTLRYRSSQF 720
QY 721 WLDEKAAALTTGQTPVCVCLPPIIHTLWQTEFDEISGNKLVTLRYARGADGGERBPRG 780
DB 721 WLDEKAAALTTGQTPVCVCLPPIIHTLWQTEFDEISGNKLVTLRYARGADGGERBPRG 780
QY 781 FGYYEQTDSHOLAQGNAPERTPPALTKWNYATGLPVIDNALSTBYWRDDQAFAGFSRFT 840
DB 781 FGYYEQTDSHOLAQGNAPERTPPALTKWNYATGLPVIDNALSTBYWRDDQAFAGFSRFT 840
QY 841 TWQDNKQVPLTPEDDNSRYWNRNALKQLLRSELYGLDSDTNKHPYTVTFEFSQVRRLLQ 900
DB 841 TWQDNKQVPLTPEDDNSRYWNRNALKQLLRSELYGLDSDTNKHPYTVTFEFSQVRRLLQ 900
QY 901 HTDSRYPVLSVSVESRNYHYERTASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQQAIPAI 960
DB 901 HTDSRYPVLSVSVESRNYHYERTASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQQAIPAI 960
QY 961 NLYPDTLPDKLLANSYDDQQRRLTYQQSSWHLHTNNTVRLGLPDSRSDIETTYGAEN 1020
DB 961 NLYPDTLPDKLLANSYDDQQRRLTYQQSSWHLHTNNTVRLGLPDSRSDIETTYGAEN 1020
QY 1021 VPAGLNLLELSDKNSLIADDKPREYLGQKTAAYTDGQNTTLPOTPTQALIAFTETTVF 1080
DB 1021 VPAGLNLLELSDKNSLIADDKPREYLGQKTAAYTDGQNTTLPOTPTQALIAFTETTVF 1080
QY 1081 NQSTLSAFNGSIPBDKLSLTLEQAGYQQTNYLFPRTGEDKVVAAHHGYTDYGTAAQFWRP 1140
DB 1081 NQSTLSAFNGSIPBDKLSLTLEQAGYQQTNYLFPRTGEDKVVAAHHGYTDYGTAAQFWRP 1140
QY 1141 QKQSNLTQKITLILWDANYCVVQTRDAAGLITTSKDYDRFLTPVQLTDINDNQHLLITL 1200
DB 1141 QKQSNLTQKITLILWDANYCVVQTRDAAGLITTSKDYDRFLTPVQLTDINDNQHLLITL 1200
QY 1201 DALGRPTILRFWGTENGKWTGYSPEKASFPSPSDVNAAIIEKPLPVAQCQVYAPESWM 1260
DB 1201 DALGRPTILRFWGTENGKWTGYSPEKASFPSPSDVNAAIIEKPLPVAQCQVYAPESWM 1260
```

```
QY 1261 PVLISQKTFNRLAEQDWOKLYNARIITEDGRICITLAYRRWVQSKAIPOLISLLNNGPRLP 1320
DB 1261 PVLISQKTFNRLAEQDWOKLYNARIITEDGRICITLAYRRWVQSKAIPOLISLLNNGPRLP 1320
QY 1321 PHSITLTTDTRDYDHPPEQOIROQVVFSDGFGRLLLQAAAARHEAGMARQNRNEDGSLIINVQHT 1380
DB 1321 PHSITLTTDTRDYDHPPEQOIROQVVFSDGFGRLLLQAAAARHEAGMARQNRNEDGSLIINVQHT 1380
QY 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDMRYVNSDSARQSEKAYADTHVYDPIGREIKV 1440
DB 1381 ENRWAVTGRTEYDNKGQPIRTYQPYFLNDMRYVNSDSARQSEKAYADTHVYDPIGREIKV 1440
QY 1441 ITAKGWFRRTILTFPWFVTNEDENDTAAAEVKVKM 1474
DB 1441 ITAKGWFRRTILTFPWFVTNEDENDTAAAEVKVKM 1474

RESULT 2
Q7N7Y2_PHOLL
ID Q7N7Y2_PHOLL PRELIMINARY; PRT; 1475 AA.
AC Q7N7Y2_
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TcdB2.
GN Name=tcdB2; OrderedLocusNames=plu0969;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boenare N., Danchin A., Kunst F.,
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571862; CAE13264.1; -; Genomic_DNA.
DR Photolista; plu0969; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR003284; Integrin alpha.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
DR Complete proteome.
SQ SEQUENCE 1475 AA; 165779 MW; 24945174CDADAD22 CRC64;

Query Match 92.2%; Score 7285.5; DB 2; Length 1475;
Best Local Similarity 91.7%; Pred. No. 0;
Matches 1352; Conservative 55; Mismatches 67; Indels 1; Gaps 1;

QY 1 MNSQDSFISITELSPKGGGAI TGMGEAL TPTGPGMAALSPLPISAGRGYAPFTLNNY 60
DB 1 MNSQDSFISITELSPKGGGAI TGMGEAL TPAAGPGMAALSPLPISAGRGYAPSLALNNY 60
QY 61 SGAGNSPGLGWCNCNMV TIRRTTHFGVPHYDETDTFLGPEGEVLVVAQPDDESTFLOGIN 120
DB 61 SGAGNTFPLGWCNCNMV TIRRTTHFGVPHYNETDTFLGPEGEVLVVAQPDDESTFLOGIN 120
QY 121 LGATFTVTGYSRLESFHSRL EYQPKTKTKTD FWL IYSPDQVHLLGKSPQARISNPSQ 180
DB 121 LGATFTVTGYSRLESFHSRL EYQPKTKTKTD FWL IYSPDQVHLLGKSPQARISNPSQ 180
QY 181 TTOTAQMLLEASVSRGEQIYYOYRAEDDTGCEADEITHHLOAQAORVLIHYVYGNRTAS 240
DB 181 TTOTAQMLLEASVSRGEQIYYOYRAEDDTGCEADEITHHLOAQAORVLIHYVYGNRTAS 240
```


Db 181 TTQTAQWLLASVPHGEIYYQYRAEDNLDCEATNEITLHPQATAQRYLHVIYGNRTAS 240
Qy 241 ETLPLDGSAPSQADWLFYLVDFYDYGERSNNLKTTPAFSTTGSWLCRODRFSRYEYGFIR 300
Db 241 ETLPLNGRAPSQADWLFYLVDFYDYGERSNNLKMPPATATGWNLCRODRFSRYEYGFALR 300
Qy 301 TRRLCROVLMYHHQAQDSKITEINGPTLVSRLLIYNDESIASTLVFVRVHGQGNV 360
Db 301 TRRLCROQLMYHCQAQDNKIKEHNGPTLVSRLLIYNDESIASTLVFVRVHGQDGT 360
Qy 361 VTLPLLELAYQDFSPRHAWQMDVLANFNAIQWQLVDLKGELGCLLYQDKGAWMYR 420
Db 361 VTLPLLELAYQDFSPHNDWQMDVLANFNAIQWQLVDLKGELGCLLYQDKGAWMYR 420
Qy 421 SAQRLGIGSDAVTWKMQPLSVIPSLQSNASLVINDINGDQLDWVITGPGLRGYSQRPD 480
Db 421 SAQRLGIGSDTWTWKIQLSVIPSLQSNASLVINDINGDQLDWVITGPGLRGYSQRPD 480
Qy 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLYANTRDGFAGKDVQV 540
Db 481 GSWTRFTPLNALPVEYTHPRAQLADLMGAGLSDLVLIGPNSVRLYANTRDGFAGKDVQV 540
Qy 541 SGTITLTPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGFOPITLPGFSQ 600
Db 541 SGTITLTPVPGADPRKLVAFSDVLGSGQAHLEVSATKVTCPNLRGRFGFOPITLPGFSQ 600
Qy 601 PATEFPAQVYLADLDSGPTDLIYVHTNRLDIFLNKSGNGFABPVTLPPEGLRFDHTC 660
Db 601 PEAIFPAQVYLADLDSGPTDLIYVHTNRLDIFLNKSGNGFABPVTLPPEGLRFDHTC 660
Qy 661 QLQWADVOGLGVASLILSVPHMSPHWRCDLTNKKPMLLNEMNNMGVHHTLRYRSSQF 720
Db 661 QLQWADVOGLGVASLILSVPHMTPHWRCDLTNKKPMLLNEMNNMGVHHTLRYRSSQF 720
Qy 721 WLDEKAALTTGQPCVCLPPIITLWQTEDEISGNKLVTLRYARGWDGRERFRG 780
Db 721 WLDEKAALVAGQTPCVCLPPIITLWQTEDEISGNKLVTLRYARGWDGRERFRG 780
Qy 781 FGYYEQDSDHQAQGNAPERTPPALTNKWTATGLFVIDNALSTEYWR-DDQAFAGFSRPF 839
Db 781 FGYYEQDSDHQAQGNASERTPPAMTKNWTATGLFVIDNTLSTEYWRGNDQAFAGFSRPF 840
Qy 840 TTWQDNKDVPLTPEDDNRVYFNALQGLRLSELYGLDSDTNKHPVYVTFEPRSQRRL 899
Db 841 TTWQDGKIDLLTPEDDNRVYFNALQGLRLSELYGLDSDTNKHPVYVTFEPRSQRRL 900
Qy 900 QHTDSRPVLMSSVYVESRNVHYRIASDPQSONITLSSDRFGQPLKOLSVOYPRROOPA 959
Db 901 QQADNQVPLVLMSSVYVESRNVHYRIASDPQSQDITLSSDULFGQPLKQVSVQYPRRKOPA 960
Qy 960 INLYPDTLPDKLANSYDDQORQLRLTYQQSSWHHLTNNTVRLVGLPDSRSDIFTTYGAE 1019
Db 961 ISPPYDPLPDELLANSYDEQORQLRLTYQQSSWHHLTDNTVRLVGLPDSRSDIFTTYEAK 1020
Qy 1020 NVPAGLNLELLSDKNSLIADKPREYVGGQKTAITDQONTPTLOTPTRQALIAFTETTV 1079
Db 1021 NVPAGLNLELLSDKNSLIADKPREYVGGQKTAITDQONTPTLOTPTRQALIAFTETTV 1080
Qy 1080 PNQSTLSAFNGSIPSDKLSLTLEAGVQOONTLYLPRFGEDKVVAAHHGYTDYGTAAQFWR 1139
Db 1081 PSQSTLSAFNGSIPSAQSTLTLEAGVQOONTLYLPRFGEDKVVAAHHGYTDYGTAAQFWR 1140
Qy 1140 PQKQNTQLTKGKILTIWDANYCVVVQTRDAAGLTTSYAKYDWRFLTPVQLTDINDNQHLIT 1199
Db 1141 PQKQNTQLTKGKILTIWDATYCVVQTQEDAGLTTSYAKYDWRFLTPVQLTDINDNQHLIM 1200
Qy 1200 LDALGRPITLRFWGTENGKMTGYSPEKASPSPDVNAALIELKPLPVAQCQVYAPESW 1259
Db 1201 LDALGRPITLRFWGTENGKMTGYSPEKASPSPDVNSAIELKPLPVAQCQVYAPESW 1260
Qy 1260 MPVLISQKTFNLAQDQWKLNYARIITEDGICITLAYERWQSOAKAPOLISLANNPRL 1319
Db 1261 MPVLISQKALNRLAECQDRQKLYNTRIITEDGRICTLAYERWHSQSQKATPQLTSLNNGPHL 1320

Qy 1320 PPHSLTTLTDRYDHDPEQOIQQVVFSDGFGRLLOAAARHEAGMARQNRNEDGSLIINVQH 1379
Db 1321 PPHSLTTLTDRYDHDPEQOIQQVVFSDGFGRLLOAAARHEAGMARQNRNEDGSLIINVQH 1380
Qy 1380 TENRWATVGTETDYNKGPITRTQPYFLNDWRYVYNSDSARQEKAYADTHVYDPIGREIK 1439
Db 1381 TENRWATVGTETDYNKGPITRTQPYFLNDWRYVYNSDSARQEKAYADTHVYDPIGREIK 1440
Qy 1440 VITAKGWRRTLTFPFWFTVNEDENDTAAEVKKVNM 1474
Db 1441 VITAKGWRRTLTFPFWFTVNEDENDTAAEVNEVNM 1475

RESULT 3

Q7N720 PHOLL
ID Q7N720 PHOLL PRELIMINARY; PRT; 1476 AA.
AC Q7N720;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TcdB1.
GN Name=tcdB1; OrderedLocusNames=plu0961;
OS Photorhabdus luminescens (subsp. laumondii)
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
RA MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Bouraux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Deros R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siguer P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunat F.,
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens.",
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571862; CAB13256.1; -; Genomic_DNA.
DR Photoclist; plu0961; -.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR00413; Integrin alpha.
DR InterPro; IPR003284; Sal SpvB.
DR Pfam; PF01839; EG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
KW Complete proteome.
SQ SEQUENCE 1476 AA; 165737. MW; 06AF0EB945B5728 CRC64;

Query Match 76.8%; Score 6065.5; DB 2; Length 1476;
Best Local Similarity 76.3%; Pred. No. 0;
Matches 1126; Conservative 120; Mismatches 221; Indels 9; Gaps 3;

Qy 1 MNSQDSITELSLPKGGGALTGMGEALTPTGPDGMAALSPLPISAGRGYAPFTLNYN 60
Db 1 MNSQTSFVAELSLPKGGGALTGMGEALTPTGPDGMAALSPLPISAGRGYSPSLTLYNS 60
Qy 61 SGAGNSPFLGWDQNVMTIRRTHTFGVPHVDEDTFLGPEGEVLVA-----DOPRES 114
Db 61 SGAGNSPFLGWDQNVMTIRRTHTFGVPHVDEDTFLGPEGEVLVAALNENGQADIRSES 120
Qy 115 TLQGINLGAFTVTGYSRLESFSLRYEYQPKTTGKTDFWLIYSPDQVHLLGKSPQAR 174
Db 121 SLQGINLGEFTVTGYSRLESFSLRYEYQPKTTGKTDFWLIYSPDQVHLLGKSPQAR 180
Qy 175 ISNPSQTTQTAQWLLASVSRGEGIIYQYRAEDDTGCEADEITHHQAQRYLHVIY 234
Db 181 ISNPLNYSQTAQWLLASVSRGEGIIYQYRAEDDTGCEADEITHHQAQRYLHVIY 240
Qy 235 GNRATSETLPGLDGSAPSQADWLFYLVDFYDYGERSNNLKTTPAFSTTGSWLCRODRFSRYE 294

Db 241 GNLTSVFTPLNGDDPLKSGWLECLVDFYGERKNSLSEIPPFKASSIMLCRQDRFSRYE 300
Qy 295 YGFEIRTRRLCRQVIMYHHLQALDSKITERHNGPTLVSRLLINLYNDESAIATSLVFRVRVGH 354
Db 301 YGFELTRRLCRQILMFHRLQTLGSGQAGDDEPALVSRKILDYDENAVISTLVSVRRIGH 360
Qy 355 EQDGNVVTLPPLLAYQDFSPRHHAHQMDVLANFNAIORWQLVLDLKGELPGLLIYQDK 414
Db 361 EDNNTVTSLPPLLAYQDFPEQKARQWQSMVDLANFNAIORWQLLDLKGEGVPGVLYQDR 420
Qy 415 GAWMYRAQRLGEIGSDAVTWKMOPLSVLPSLQNASLVLDINGDGLDWDVITGPGLRGY 474
Db 421 NGWYRQAQQAQGEEMWAVTGWKQLLPITPALQDNASLMDINGDGLDWDVITGPGLRGY 480
Qy 475 HSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLDLVIGPKSVRLYANTRDQFAK 534
Db 481 HSQHPDGSWTRFTPLDALPIEYSHPRQAQLADLMGAGLSLDLVIGPKSVRLYANNRDGFTQ 540
Qy 535 GKDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLYEVSATKVTCPNLRGRGFGQPIIT 594
Db 541 GRDVVQSGDITLVPFGADARKLVAFSVLGSGQAHLYEVSATQVTCWPNLHGGRFGQPIIT 600
Qy 595 LPGFSQATFBNPAQVYLAIDLGGSGPTDLIVYHTNRDLIFLNKSGNGFAEPVTLRPPGGL 654
Db 601 LPGFSQADNFNPRVHLADLGGSPADLIVYHTDRLEIFKNESGNSFAKPTLRFPDGL 660
Qy 714 RFDHCTQLQADVQGLGVASLILSVPHMSPHHWRCDLTNMPMLNEMNNMNGVHHHTLRY 714
Db 661 RFDHCTQLQADVQGLGVASLILSVPHMAHHWRCDLTNAKPMLLSEMMNNMGAHHTLRY 720
Qy 774 RSSQFWLDEKAAALTTGQTPVCVLPPIHTLWQTETEDISGNKLVITLRYARGAWDR 774
Db 721 RSSQFWLDDKAAALATGQTPVCVLPFPVHTLWQTETEDISGNKLVITLRYARGAWDR 780
Qy 833 EREPRGFGYEOEDSHOLAQGNAPERTPPALTKNWYATGLPVIDNALSTEWYR-DOQAF 833
Db 781 EREPRGFGYEOEDSHOLAQGNAPERTPPALTKNWYATGPEVDNTLSAGYWRGDKAFT 840
Qy 834 GFSRFTTQDNKDVPILTPEDDNRYSYFNFRALKQGLRLSELYGLDDSTNKHVPYVTVTEFR 893
Db 841 GFTPRFRWKEGKDVPAATPENNDDLYFNFRALKQGLRLSELYGLDDSEQQNPVTVTESR 900
Qy 894 SOVRRLHTDSRYDVLNHSVVSERYHYERIASDPQCSQNTILSSDFGQPLKOLSVOYP 953
Db 901 PQVRQLQDGTASVPLWASVVENRSYHYERIIIGPQCNQDITLSSDQFGQPLKQVSMQYP 960
Qy 954 RRQOPALNLYPDTLPDCKLANSYDDQORQLRLTYQSSWHHLTNNTVRLGLPDPSTRSDI 1013
Db 961 RNRQPTNPPYDITLPDILPASSYDDQQLLRLTYQSSWHHLTGNELRVLGLPDPGTRSDA 1020
Qy 1014 FTYGAENVAGGLNLELLSDKNLSLIADKPREYLGQOKTAYTDQNTPTIQTPTROALIA 1073
Db 1021 FTYDAKQVPVGLNLEALCAENSLIADKPREYLNQORTFTYTDGKNQAPLEIPTROALIA 1080
Qy 1074 FTEFTVFNQSTLSAFNGSIPSDKLSLTLEAGYQOTWYLPRTCTEDKVVAAHGYTDYGT 1133
Db 1081 FTETAVALTESLSAFDGGITPDELPGLITQAGYQOEYLPFRTEGNKVVAAHGYTDYGT 1140
Qy 1134 AAQFWRPQKQSNTOGLTKITLIWDANTCVVYQTRDAAGLTTSYAKYDWRFLTPVQLTDIND 1193
Db 1141 EAQFWRPVQAQNTLLTGKTLTLOWDTHYCVITQTDAGLTVLANYDWRFLTPVQLTDIND 1200
Qy 1194 NQHLITLDALGRPTLRFWGTENGKMTGYSSPEKASFPSPSDVNAALELKKPLPVAOCQV 1253
Db 1201 NVHLITLDALGRPVQTFWGTENGKMTGYSSPEKPFSPPTDINTAIALTGPLPVAQCLV 1260
Qy 1254 YAPESWMPVLSQKTFNLAQDMQKLYNARIITEDGRICTLAYERWQSOXAIPLQLISLL 1313
Db 1261 YAPDSWMPVLSQKTFNLTUEEQTLRLDRIITEDWRIICALRRWLSQSOASTPLVNL 1320
Qy 1314 NNGPRLPSPHSLTLTDRYDHDPEQOIRQQVWFSDFGRLGLQAAHBAQMARQNEGDGSL 1373

Db 1321 TNSIGLPPHNLTLTPDRYDRDSGQIHOQVAFSDGFGRLQLQASVREHAGEAWQRNQGSL 1380
Qy 1374 IINVOHTENRWAVTGRTEYDNKQOPRTYQPYFLNDWRVYVNSDSARQEKAYADTHYDYP 1433
Db 1381 VTKMEDTKTAVTGRTEYDNKQGTIRTYQPYFLNDWRVYVSDSAR--KGAYADTHYDYP 1438
Qy 1434 IGREIKVITAKGFRRTLTPTWFTVNEDENDTAAEV 1469
Db 1439 IGREIRVITAKGWLROSQYFFWFTVSEDENDTAAV 1474
RESULT 4
Q3EP6 PHOLU
ID Q3EP6 PHOLU PRELIMINARY; PRT; 1476 AA.
AC Q3EP6;
DT 01-DEC-2001 (TEmBLrel. 19, Created)
DT 01-DEC-2001 (TEmBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TEmBLrel. 26, Last annotation update)
DE TcdB1.
GN Name=tcdB1;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA french-Constant R.H.;
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RT "Genomic islands in Photorhabdus.";
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AAL18487.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR003284; Sal_SpVb.
DR Pfam; PF01839; FG-GAP; 2.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSVPBPROT.
SQ SEQUENCE 1476 AA; 165138 MW; 8E6AC3D109911995 CRC64;
Query Match 76.3%; Score 6031.5; DB 2; Length 1476;
Best Local Similarity 75.9%; Pred. No. 0;
Matches 1120; Conservative 122; Mismatches 224; Indels 9; Gaps 3;
Qy 1 MNSQDSFISITELSLPKGGGALTGMGEALTTPGPGMAALSILPLPISAGRGYAPAFITLYN 60
Db 1 MNSQSFVITELSLPKGGGALTGMGEALTTPGPGMAALSILPLPISAGRGYAPAFITLYN 60
Qy 61 SGAGNSPFLGWDGNCVMTIRRTTHFGVPHYDETDITFLGPEGEVLVVA-----DQPRDES 114
Db 61 SGTGNSPFLGWDGNCVMAIRRTSTGTVPNTDITFLGPEGEVLVVALEAGQADIRSES 120
Qy 115 TLOGINLGATFTVTVSRLESFHSRLEYWQPKTKTDFWLIYSPDQVHLLGKSPQAR 174
Db 121 SLOGINLGATFTVTVSRLESFHSRLEYWQPKTKTDFWLIYSPDQVHLLGKSPQAR 180
Qy 175 ISNPSQTTQTAQMLLEASVSSRGEQIYYQYRAEDDTGCEADEITHHQLATAQRYLHIVY 234
Db 181 ISNPLNVNQTAAQMLLEASISHSSEQIYYQYRAEDDEACETDELAHAHPSATVQRYLQVHY 240
Qy 235 GNTASFTLPGDGSAPSQADWLFYLVDFYGERSNLKTTPAPSTTGSWLCRQDRFSRYE 294
Db 241 GNLTSADVFTPLNGDDPLKSGWMLFCLVDFYGERKNSLSEMPLEKATGNWLCRQDRFSRYE 300

Qy	61	SGAGNSPGLGWCDCNVMTIRRRTHFGVPHYDETTTFI GPGSEVLVA-----DQPRDES	114
Db	61	SRSGNGPFGIGWGIGGAAVQRRTNGAFTYDDTDEFTGPDGEVLVPALTAAGTQEARQAT	120
Qy	115	TLOGINLGAFTVTVGYSRSLSHESFRLEYWPKTKTKTDFWLIVSPDQGVHLLGKSPQAR	174
Db	121	SLGINPGGSFNVQYASRTSGLSRSLERWLPADETETEFVWLTVPDQVALLGRNAQAR	180
Qy	175	ISNPSOTTQATQWILLESVSRSGEQIYYQYRAEDDTGCEADEITHHLOATAQRYLHIVYY	234
Db	181	ISNPTAPTQATVWLMESSVSLTGQMYYYQYRAEDDDGCDRAERDAHPQAGQORYPVAWY	240
Qy	235	GNRFTASSTLPGLDGASASQADWLPIYIPVDYGGERSNNLKTPPAPSTTGS--WLCRQDRFSR	292
Db	241	GNRQARTLPAL-VSTPESMDSWLFILPVDYGGERSVLSSEAPAMQTPGSGMWLCRQDCFSG	299
Qy	293	YEFGEFTRRLCRQVLMYHHLQALDSKI TEHNQPTLVSRLLILNYDESATLTVFVRRV	352
Db	300	YEFGFNTRRLCRQVLMFYHVLVAGSSGANDAPALISRLLLDYRESPSLLENWHQV	359
Qy	353	GHEODGNVVTLPPELAYIQPFSPRHHAHQPMQVLANFNAIQRWQLYDLKGEGLGLLYQ	412
Db	360	AYESDGTSCALPALALGQWTFPTPTLSAQWTRDDMGKLSLLQPYQYDLNAGEVGVILYQ	419
Qy	413	DKGAWWTRSAQRIGEISDAVWWEKMQPLSVIPSLQSNASIVDINGQGLDWITGBCLR	472
Db	420	DSGAWWYREPRVQSGDDPDVWVGAAALPTMPALHNSGTLADLNGRLEWVVTAPGVA	479
Qy	473	GYSORPDGSWTRFPTLNALPVEYTHPRAOLADLMGAGLSDLVLIGPKSVRLVANTRDGF	532
Db	480	GMYDRTFGRDMLHFTPLSALPVEYAHKPAVLADILGAGLTDWVLIGRPSVRLYSGKNDGW	539
Qy	533	AKGKDVQSGDITLPPVGPADPRKLVAFSDVLGSGQAHLEVSAKTVCWPNLGRGRGPQ	592
Db	540	NKGETVQOTERLTLPVPGVDPRTLIVAFSDMAGSQOHLTEVRANGVRYVWPNLGHGRGPQ	599
Qy	593	ITLPGFSOPATEENPAOVLADLQSGPTDLIVYHTRLDIPLNKGSGFAEPVTLRFFE	652
Db	600	VNI PGFSQVTTTFNPDQILQADTDGSGTDTLIVANSRLVIFYNQSGNYFAEBHTLLLPK	659
Qy	653	GLRFDHTCOLQADVQGLGVASLILSVPHMSPHHWRCDLTNMRKFWLLNNNNNGVHHTL	712
Db	660	GVRDTRCSLQVADIQGLGVPSLLTTPHVAPHWVCHLSADKFWLLNGNNWGAHAL	719
Qy	713	RYRSSQFWEIDKAAALTTGQTPVCYLPPIHILWQTEDETSIGNKLVTLTRYARGAWD	772
Db	720	HYRSSVQFWEIDKAEALAAAGSSPACYLPFTLHTLWRSVWQDEITGNRLVSDVLYRHGVWD	779
Qy	773	GRREPRGFGVQECTDSHOLA-QQNAPERTPPALTKNWYATGLPVIDINALSTEVRRD-Q	830
Db	780	GQREPRGFGFVEIRUDTTLASQATATELSMPSSVRNWIATGVPADVDERLPETWQNDA	839
Qy	831	AFAGFSPRFTTQWQNDKVPLTPEDDNGRYWYFNALKGOLLSELVGLDDSTNKHVPYTVT	890
Db	840	AFADFAFTVSGSEDQTYTP-DDSKTFWLQRALKGLILRSELYGADGSSQADIPIYSVT	898
Qy	891	EFBSQVRRLQHTDSRYPVLMSSVVESSNYHYERIASDPQCSQNTLSSDRFGQPLKQLSV	950
Db	899	ESRPQV-RLVEANGDYVPMWPMGAESRTSVYRYHNDPQOQQOAVLLSDEYGFPLRQSV	957
Qy	951	QYPRQCPQANLYPDTLPDKLLANSYDQQRQLRLTYQQSSWHHLTNNTVR--VLGLPDS	1008
Db	958	NYPRPSPANDPYPASLPATLFPANSYDEQQIILNLGLQQSSAHLVSLSGHWLLGLAEA	1017
Qy	1009	TRSDIFTYGAENVPAGLANLELSDKNLSLIADKPREYLGQOKTAYTDGQNTTLPLOTPTTR	1068
Db	1018	SRDDVPYTSADNVPEGGJLLEHLLAPESLSVSDSQVGTLAGOQQQWYVLDSDQVATVAAPPL	1077
Qy	1069	QALIAFTTTFVFNQSTLSAFNGSIPSKLSTLEOAGYQOYNTLYFP--R'GEDKVVVAHH	1126
Db	1078	PPKVAFLETAVLDEGMWVSSLAAYIVDSDH----LEQAGYRSGYLFPGREABQALMTQCC	1133

Qy	1127	GYTDYGTAAQFWPBPQKOSNTQLTGKLTILWDANYCVVVQTRDDAAGLTTSKAYDWRFLTPV	11886
Db	1134	GYVYGAEBFWLPLSFLDRSMLTGPVTVTRDAYDVCITQWQDAAGVTTTADYDWRFLTPV	11993
Qy	1187	QLTDINDNQHLLITLDALGRBITLRFWGTENGKMTGYSSPEKASFSPSDVNAALCLKPL	1246
Db	1194	RVTDPNDNQSQVTLDALGRVTTLRFWGTENGITATGSD---	1350
Qy	1247	PVAQCOVYAPESWMPVLSQKTFNRLAEQWQKLNYARIITDGRICTLAYRRVWQSKAI	1306
Db	1251	PVAQCLVYVTDSW-----GDDNEK-----	1270
Qy	1307	PQLISLNNNGRLPPHSLTLITTDYDHPPEQQIRQQVVFSDGFRLLQLAARHEAGWARQ	1366
Db	1271	-----MPPHVVVLATDRYSDTCQQVQRQVTFSDGFRGLQSATRQAEGNAWQ	1318
Qy	1367	RNBDGSLI-----INVQHTENRWAVTGRTEYDNKGQPIRTTQPIFLDNRVTVSND SAR	1419
Db	1319	RGRDGLKLVITASDGLPVTVTA--TNFRWAVTGRAEYDNKGLPVRVQPYFLOSQVYVSDSAR	1377
Qy	1420	QOEKAYADTHVYDPIGREIKVITAKGFWRFTLTPHPFTVNEDENDT	1465
Db	1378	Q--DLVADTHFYDPTAREQVITAKGERQVLVTPWFVSEDENDT	1421

RESULT 9

```

06QDZ4_YERPS
RESULT_9
AC Q6QDZ4_YERPS PRELIMINARY; PRT; 1489 AA.
ID O6QDZ4; Q665G5;
AC Q6QDZ4; Q665G5;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Y0185-like protein (insecticidal toxin complex).
GN Name83ac; Ordered locus names=YPTB3553; ORFNames=YPP3678;
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OC NCBI_TaxID=633;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=IP32953;
RA Bernardes Pinheiro V.B., Ellar D.J.;
RL Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.
[2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=IP23253 / Serotype I;
RX PubMed=15358858; DOI=10.1073/pnas.0404012101;
RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,
RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,
RA Rubakar R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C.,
RA Simonet M., Chenal-Francisque V., Souza B., Dacheux D., Elliott J.M.,
RA Darbisee A., Hauser L.J., Garcia E.;
RT "Insights into the evolution of Yersinia pestis through whole-genome
RT comparison with Yersinia pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004) .
DR EMBL; AY545599; AA566065.1; -; Genomic DNA.
DR EMBL; BX936398; CAH22791.1; -; Genomic DNA.
DR GO; GO:0005737; Cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin_alpha.
DR InterPro; IPR000408; Reg_chr_Condens.
DR InterPro; IPR003284; Sal_SpVb.
DR Pfam; PF01839; FG-GAP; 1.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSVPBPROT.
DR PROSITE; PS00626; RCCI_2; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 1489 AA; 167592 MW; CEASB7C0A04B7B44 CRC64;

```

Query Match 46.0%; Score 3638; DB 2; Length 1489;
Best Local Similarity 50.1%; Pred. No. 6.3e-223;
Matches 756; Conservative 214; Mismatches 476; Indels 62; Gaps 26;


```
DR PRINTS; PR01341; SALSVPBPR0T.
DR PROSITE; PS00626; RCC1_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1496 AA; 168681 MW; EB3E8E9BC8917C2 CRC64;

Query Match      45.2%; Score 3569.5; DB 2; Length 1496;
Best Local Similarity 49.3%; Pred. No. 1.5e-218;
Matches 748; Conservative 211; Mismatches 486; Indels 71; Gaps 29;

Qy 1 MQNS-QPFSITSLPDKGGGAIITGMGEALTTGPDGMAALSPLPIPIAGRGYAPAFITLY 59
Dy 1 MENSQKVAVAPVAVSLPDKGGGAIITGMGSLGPIGSGMATLTPLPIAGRGYAPSLTLY 60

Qy 60 NSGAGNSPGLGDCNVMTTRRTTFGVPHYDETDTLGPEGEVLVADQPD--ESTLQ 117
Dy 61 SSGSGNGPFLGQWLGQWTAIRRRNAQVPRYDEYDEFLAPNGEVMVVAADPQGSIRTEQ 120

Qy 118 GINLGATFTVTVGYSRSLSHESFSLLEYWQPKT-TGKTDFWLIYS PDGQVHLIGKSPQARIS 176
Dy 121 SLN-GEQFSVIRYLPRIEGNPHRIEYWRPTNNSQAPFWLVHSSDGQKHCGLGYSAAARIA 179

Qy 177 NPSOTTOTAOWLLEASVSSRGEIYYQYRAEDDTGCEADEI-----THHQAATQRYLHI 231
Dy 180 DPLPHEHIAEWLLESVLSGEHIGYQYQADEQGIIDEPISYKAEKQNHVPAASAQRYLKR 239

Qy 232 VYGNRTASLTPLGLDGSAPSQADWLFYLVFDYGERSNLTKTPAFSTTSGWLCRODRFS 291
Dy 240 VYGNRAQAEYLYCLT-QQPAPTSWFLSLFDHGEYSNIAEQVPIVIGKSWNRQDAFS 298

Qy 292 RYEGFETIRRLCRQVLMYHHLQALDSKITERNHGPTLVSLRLIINYDESIASTLVFVR 351
Dy 299 HFNTGFVETRLCCQVLMYHNLKALGDEPDQAQ-TLVSLRLHYQHDATVATQVGCQ 357

Qy 352 VGHQDGNVVTLPLELAYQDFSPRHHAHQPMQVLANFNAIQRLVDLKGEGLPGLLY 411
Dy 358 LAHEPDGTGRSLPLEFDYQDFSTRDALGWOPLTDWAEFN--YQYQWVDLNGEGMPGMLY 415

Qy 412 QDKGAWYRSQRLGEITG-SDAVTWKQWPLSVIPSLQSNASLVINDGOLDWVITGPG 470
Dy 416 QDSGHWIRYRPVR--QPTAGDITFGAAQRLPSIPAMKNAMLMINDGDKLQWVISOQP 473

Qy 471 LRGYH-----SORPDGSWTRFTPLNALPVETTHPRAQLADLMGAGLSDLVLIQKSVRL 524
Dy 474 LAGYFSRDPDLRSRDPDLSTQFIPLSITLPAEYFHPQAQLVDLAGSLDLALVGPKSVRV 533

Qy 525 YANTRDGFAGKGVVQSGDITLTPVGADPRKLVAFSDVLGSGQAHVSEVATKVTCPNL 584
Dy 534 YTNLCDSPFAAATQVADDDITLPLPGVHFTLVAFSDVMGSGQQLVRIHNSVTCWPNL 593

Qy 585 GRGFGOPITLPGFSQAPATEENPAQVVLADLDGSGPTDLIVVHTNRLDIPLKNSGNGFAE 644
Dy 594 GHGFGHPLSLPGFNGPVEQFNPLAIYADLDGSGTIDLIVATTSLQLIYRNQSGNPAE 653

Qy 645 PVTLRFPEGLRFDHTCOLQADVOGLGVASLILSVPHMSPHHWRCDLTNMKFWLLNEMNN 704
Dy 654 PLATLPTGIRFNSCQLSLADIQGLGVASIMLSVPHPTTQHWYDFVASKPYLLCTNN 713

Qy 705 NMGVHHTLRVYSSSQFMLDEKAAALTTGQTPVCVLPPIHTLMQTEDEISGNKLVVTL 764
Dy 714 NMGAESQLLYRVSQFVWLDKAAQAQKGRSLACQLPFPPIHLLAQTTQDEITGNSLQTA 773

Qy 765 RYARGANDGRRERFEGYVETQSHQACQNAQNPRTTPALTKWYATGLPVIDNALSTE 824
Dy 774 RYFHGFYDGVQREBSFGFRVDTLDTDTSAQSAARAAPTAKSSRWFTHTRAGNETLMQSE 833

Qy 825 YWR-DDQAFAGFSRFTTQDNK--DVPLTPTEDDNRYSWFRNALKGQLLRSELYGLDDST 881
Dy 834 YWQGDQAYSLLPTFLTKFINNTQGDLLSELDDNQTFWLHRLKGLSLRSELYGLDDSE 893

Qy 882 NKHVPTVTEPRSQVRLQHTDS--RYPVLMSVSVESNHYHYRIASDPQCSQMITLSSD 939
Dy 894 LATQPSYNSRYQVRQTQSSADGTSFSPVALPMVLEQLSYHYERIAQDPQCSQOQVILRCN 953
```

```
Qy 940 RFQOPLKQLSVQYPRROQPAINLYPDPDLPDKLANSYDDQORQLRLTYQSSWHHLTNT 999
Dy 954 EYGHPLHSVINYPRRDKARISPY-SWLAKHEHWSHFEQOQLRITESQSYHHEISDK 1012

Qy 1000 VRVLGLPDSRSDIFTTYGAENVPAAGLNLELLSDKNSLIADDPREYLGQOKTAYTQGN 1059
Dy 1013 FVVLGLPAGQSDVLITYPDNFVPTAGIHWBELQOPEGLLGTAKERTTFAQQQVFT--SD 1070

Qy 1060 TTPLQTPROALTAFTETTVFNOSTLSAFNGSPSPDKSLTTLLEQAGYQQTNYLPRTGED 1119
Dy 1071 TIP-----GLWAYSQQAEDDQTLVALDELUPANERKQOLI KAGIOIAPRLPARTGET 1123

Qy 1120 KVVVAHGYTDYGTAAQFWRPQKQSNLTQLGKTIILWIDANYCVVVQTRDAAGLTTSKYD 1179
Dy 1124 DIWVAQSGFTDYGDASRFYRPIQSRSQTQLVGKTILEWDATCCAVSDIILLADYSITHAEYD 1183

Qy 1180 WRFTPTVOLTDINDNOHLITLDALGRPITIRFWGTE-----NGKM--TGYSSEPKASPSPP 1233
Dy 1184 YRFITPYLLDINDNOHYIELDALGRVTSRFACTEIDPQTNKVIETGFTPSIAEQPSAP 1243

Qy 1234 SDVNAATELKPK-LPVAQCOVYAPESWMPVLSOKTFNRLAEQDW-----QKLYNARI 1284
Dy 1244 NSVDKALSLENTIRIPVAQFSYQFQSWMISLQDDI-----EIWRANNITPEYLFQNH 1298

Qy 1285 ITEDGRICTLAYRRW-VQSOKAIPQLISL-LNNGPRLPPLHSLTLTDRYDHDPE-QQIRQ 1341
Dy 1299 LIDNYLCPLALRWGFGQNNLLITGEGVLTKNPMRQPHLLTVVDNYFSASEPQHQHQ 1358

Qy 1342 QVVFSDGFGRLQAAARHEAGMARQNEGSLIN-----VQ-HTENRWAVTGTEYDNK 1395
Dy 1359 TLAFSDCGFRVLLSARVETGSPSYDFPENGLLVDDKGNLVQLEVDQRMWVSGSTEYDNK 1418

Qy 1396 GQPIRTYQYFLND--WRVYVNSDARQEKAYADTHVVDPIGREIKVITAKGWFRRTLT 1453
Dy 1419 GLPRRYQYQYFFDNWIMLYIANN--RTLKEAYADTHYDPLGREIKVITAKGYLURTHYF 1476

Qy 1454 PWFTVNEDENDTAAEV 1469
Dy 1477 PWFVISEDNDTASEI 1492

RESULT 12
Q8DIP6 YERPE PRELIMINARY; PRT; 1516 AA.
AC Q8DIP6;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative toxin subunit.
GN Name=tcaC1; OrderedLocusNames=y0185;
OS Yersinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
[1]
RP NUCLEOTIDE SEQUENCE
RC STRAIN=KIM5/Biovar Mediaevalis;
RX MEDLINE=22137863; PubMed=12142430;
RX DOI=10.1128/JB.184.16.4601-4611.2002;
RA Deng W., Burland V., Plunkett G. III, Boutin A., Mayhew G.F., Liss P.,
RA Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C.,
RA Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V.,
RA Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R.,
RA Perry R.D.;
RT "Genome sequence of Yersinia pestis KIM."
RL Bacteriol. 184:4601-4611(2002).
DR EMBL; AE013618; AAM83779.1; -; Genomic_DNA.
DR GO; GO:0005737; C:cytoplasm; IEA.
DR GO; GO:0008305; C:integrin complex; IEA.
DR GO; GO:0007160; P:cell-matrix adhesion; IEA.
DR InterPro; IPR000413; Integrin alpha.
DR InterPro; IPR00408; Reg_chromodomain.
DR InterPro; IPR003284; Sal_SpVb.
```

DR	Pfam; PF01839; EG-GAP; 1.	
DR	Pfam; PF03534; SpvB; 1.	
DR	PRINTS; PR01341; SALSPVBPROT.	
DR	PROSITE; PS00626; R0C1.2; UNKNOWN 1.	
SQ	SEQUENCE 1516 AA; 170748 MW; 40DE58460A811CA7 CRC64;	
Query Match 45.2%; Score 3569.5; DB 2; Length 1516;		
Best Local Similarity 49.3%; Pred. No. 1.5e-218;		
Matches 748; Conservative 211; Mismatches 486; Indels 71; Gaps 29		
Qy	1	MONS-QDPSITELSLPKGGGAIATGMEALTTPDGMALSLPLPISAGRGVAPAFITLNY 59
Db	21	MENSKQQVAVAPLSLPKGGGAIATGMSDSLGPIGSGMATLTLPLPISAGRGVAPSLTILSY 80
Qy	60	NSGAGNSPFGLGWDCNVMTIRRTTHFGVPHYDETDTFLGPEGEVLVADQPRD--ESTLQ 117
Db	81	SSGSGNGPFGLGWGLTGWAIRRTNAQVPRYDEYDEFAPNGEVMVVAADFGSIERTEQ 140
Qy	118	GINIGATFTVTGYRSRLSHSRSLRYWPKT-TGKTDFWLIYSPDQGVHLLKGSQPARIS 176
Db	141	SLN-GEOSVTRYLPRIEENFRHRIEYWRPRTNNSQAPFWLVHSSDQGHCLGYSAARIA 199
Qy	177	NPSQTTQTAOWLLLEASVSSRGEQIYYOYRAEDDGTCEADEI-----THLQATQRYLHI 231
Db	200	DPLHEPIHAEWLLESVLSGEHIGYQQAEDQIDEPISYKAEQNHQAASQRYLKR 259
Qy	232	VYGNRTASETLPLGLDGSAPQADWLFLVFDYGYERSNNLKTTPAFSTTGSWLRCQDRFS 291
Db	260	VYGNRQAAYELYCIT-QQPAPTSLWFSLIIDHGEYSNIAEQVPVIKKGSMNFRQDAFS 318
Qy	292	RYEYGEIRTRRLCRQVMYHHLQALDSKITHEHNGPTLVSRLLNYDESAIASTLVFVRR 351
Db	319	HFNYGFEVTRRLCQQVLMYHNLALKGDDEPDAQA-TLVSRRLHYQHDAVATQVLGCGQ 377
Qy	352	VGHEDQGNVTVLPLELAYQDPSRRHHAHWQPMDVLANFNAIQRWLVDLAGEGLPGLLY 411
Db	378	LAHEPDGTGKSLPPLFDFYQDFTRDALGWQPLTDWAEFN--YYQWVDLNGEGMPGMLY 435
Qy	412	QDKGAWYRSARQLGEIG-SDAVTWKMQPLSVIPSLQSNASLVDINDGQGLDWITGPG 470
Db	436	QDSGHWLYRPVR--QPCTADGITFGAQRPLPSLPAMRENAMLMIDNGDKLDWVISQPG 493
Qy	471	LRGYH-----SQRPDGSWTRFTPLNALPVYTHPRAQLADLMAGAGLSDLVLIQPKSVRL 524
Db	494	LAGYFSRDPDLSDRPDLSTWQFIPLSTLPAEYFHPQALVDLAGSGLSDALVGPSPKSVR 553
Qy	525	YANTRDGFAGKQDVQSGDITLVPFGADPRKLAVSFVLSGSOAHLVEVSATKTCWNL 584
Db	554	YTNLCDSFAAAQTVAQDDITLPLFGVHFTLFAFSDVMGSGQQHVLVIRHNSVTCWNL 613
Qy	585	GRGFGGPIITLPGFSQATEFNPQAVYLADLDGSGPTDLIVYHTRLDLFLNKSNGFAE 644
Db	614	GHGRFGHPLSLPGNGQPVQEPNPALYIADIDGGSTDLIYATTSQLLIYRNQSGNRF 673
Qy	645	PVTLRFPFGLRFDHTCOLQMAADVQGLGVASILSVPHMSPHHWRCDLTNMPKPLLINMNN 704
Db	674	PLAIALPTGIRFDNSCQLSLADIQGLGVASIMLSVPHPHTQHWRYDVFVASKPYLLCTINN 733
Qy	705	NMGVHHTLYRYSQFMDEKAAALTTGQTPVCVYLPPIPIHTLMQTEDEDEISGNKLVTTL 764
Db	734	NMGABSQLLYRASSVQFMDEKAAQKQGRSLACQLPFIHLIAQTQTFDEITGNSLSOTA 793
Qy	765	RYARGAWDGRERERFGYVEQTDSHQLAQNAPERTTPALTKNWYATGLPVIDNALSTE 824
Db	794	RYFHGFDYGVOREFSGFRVDLTLDTTSAQGSAAERTAPTSSRWFFHTRAGNETLMQSE 853
Qy	825	YWR-DDQAFAGFSRFTTWQDNK--DVPLTPEDDONSRYWFNRALKGQLLRSELVGLDDST 881
Db	854	YWQGDQAYSLLPRLTKFINNTQGEDLLSELDDNQTFWLHRLKGSLLRSELVGLDDSE 913
Qy	882	NKHVPYTVTFRSQVRRLQHTDS--RYPLVNSSVVSERNHYERIASDPQCSQMITLSSD 939
Db	914	LATQPSYNSRSYQVROIQSSADGTSPPVALPMVLEQSYHYERIAQDPQCSQIVLRCN 973

```
Query Match      42.4%; Score 3350.5; DB 2; Length 1401;
Best Local Similarity 48.6%; Pred. No. 1.3e-204; Mismatches 474; Indels 53; Gaps 21;
Matches 685; Conservative 197;

Qy 1 MNSQDSITELSPKGGGALTGMGEALTPTGPGDMAALSPLPISAGRGYAPAFITLNYN 60
Db 1 MGSSTPLKLETPSPSGGSLKNGEALNVAEGGASFSPLPISVGRGLVPVLSLNS 60

Qy 61 SGAGNSPFLGWDGNVMTIRRTTHFGVPHYDETDTFLGPEGEVL-VVAD---QP-RDES 114
Db 61 STAGNSFGMGWQCGVFSISRTAKGVPHYTGQDEVILGPDGEVLSIVPDSQOGEQRTAT 120

Qy 115 TLOGINLGATFTVGYSRLESFSLRYMQP--KTTCKTDFWLLYDGGGOVHLGKSPQ 172
Db 121 SLTGVTIQTPTVTHYQSRVAEKIVRLEHWPOQRREETSFWLFTADGLVHLFGKHH 180

Qy 173 ARINSPSTTTQAOWLEASVSSRGEQIYYOYRAEDDDTGCEADBITHH--LQATAORYLH 230
Db 181 ARIADPQDETRIAELWMEETVTHGEHIYYHYRAEDDLDCEHELAQHSVLRPTLSWQ-- 238

Qy 231 IVYGNRTASTLPGCLGSAPOADWLPYLVDYGERSNLKTTPAPSTTGS----- 282
Db 239 -VOYGNTOPTAFPAVKSIGIPVDNDWLFHLVDFYGERLSSLSNVPEFVNSENNVS 297

Qy 283 --WLCRODRESRYEYGEIRTRRLCROVLYHHLQALDSKITEHNGPTLVSRLLILNDES 340
Db 298 EKWCRDPDSFSGEYGEIRTRRLCROVLMFQKALAGEKVAETPALVSRLLLDYDLN 357

Qy 341 AIASTLVFVRVGHQGNVVTLPPELAYQDFSPRHHAHQMPMDVLNANAIQRWOLVD 400
Db 358 NKVSLQTARLAHETDGTPTWMSPLENDYQRVNHGVNLWQSPQLEKNTLPQYQLVD 417

Qy 401 LKGEGLPGLLYQD-KGAWWYSAQR-LGEIGSDAVTWKMQPLSVIPSLQSNASLVING 458
Db 418 LYGEIGPALLYQDTQKAWWYAPVRDI-TAEGTNAVTVYEEAKPLPHIDPAQBSAMLLDING 477

Qy 459 DQOLDWITGCIAGYHSQRDGSWTRFTPLNALPVEYTHPRAQLADMGAGLSDLVLIG 518
Db 478 DGRLDWITASGLGYHTMSPEGETPTPIPLSAVMEYFHFQAKLADGAGLPDLALIG 537

Qy 519 PKSVRLYANTRDGPAKGDVQVSGDITLTPVPGADPRKLVAFSVGLSGQAHLVSVATKV 578
Db 538 PMSVVMNSNPAGWDRAQDVHLNKPPLVPGKKNHILVAFSDMTGSGQSHLVEVTANSV 597

Qy 579 TCWNLGRGRGQPIITLPGSQPATEFNPAQVYLADLDSGPTDLIYVHTNRLDIFLNS 638
Db 598 RYWNELGHGKFGPELMTITGQITGKRLTPTDCIWTYMAQAPPDFIYARNITYLELYANES 657

Qy 639 GNGFAEPVTLRFPEGLRFDHTCOLQADVQGLGVASLILSVPHMSPHHWRCDLTNMPWL 698
Db 658 GNHSAEPQRIIDLPDGVFPDDTCRIQIADTQGLGTASILLITPHMKVQHWRLDMTIFPWL 717

Qy 699 LNMNNNGVHTLRYSSSQFWLDEKAAALTTGQTPVCYLPFIPIHTLMQTEDEISGN 758
Db 718 LNAVNNMGTTLLYRSSAQFWLDEKLOQASGSMVTVSYLPFPVHLWRTVLEISGN 777

Qy 759 KLVTTLRYAGAWDGRBREFFGFGVGEQTDHQIA---QGNAPERTTPALTKNWTATGLP 815
Db 778 RLTSYHYSHGAWDGLEREFGRGRTVQTQDIDRSASATQGTTHABPPAPSRPTVNNYGTGR 837

Qy 816 VIDNALSTYWR-DDOAPAGSPRFTTWQDNK--DVPLTPDDNSRYWFRNALKGOLLRS 872
Db 838 EVDILLPTEYWGQDQAFPHFTTPRTPRYDEKSGGDMVTTP-SEQEYVHLHAKGQRLRS 896

Qy 873 ELYGLDSTNKHVPVTVTFERSQVRRLOHTDSRPVLMSSVSVESRNYHYERIASDPQCSQ 932
Db 897 ELYGDDDSILAGTYSVDESRTQVRLPVMVSDPAVLVSAESRQYRYERVAITDPQCSQ 956

Qy 933 NITLSSDRFGQPLKOLSVOYPRROOPAINLYPDTLPDKLLANSYDDQQRQLRLTYQSSW 992
Db 957 KIVLKSALGFPQDNLEIAYRRRPQPEFSPYPTLFTSSPDEQMQFLRLTRORFSY 1016
```

```
993 HHLT-NNTVRVLGLPDSRSDIFTYGAENVPAAGLINLELLS--DKNSLIADDKPREYLGQ 1049
1017 HHLNDDNTWITGLMDTISRSDARIYQADKVPDGGFSLWFSATGAGALLLPDAADYLGH 1076
1050 QKTAITDQONTPTLOTPTROALIAFTTTVPNOSTLSAFNGSIPSDKSLSTTLEAQVQOT 1109
1077 QRVAYTC-----PREQPAIPLVAYIETABFDSRLAAAFBEVMDQEQLTKQLNDAGWNTA 1131
1110 NYLFPRTGEDKVMVAHHGYTDYGTAAQFMRPQKSNLTQKITLIINDANYCVVVQTRDA 1169
1132 KVPFSEKTDTHVWVGQKEFTEYAGADGFYRPLETKTGKTTVTWDSHYCVVITATEDA 1191
1170 AGLTSAKYDWRFTPTVQLTDINDNQHLITLDALGRPITLRFWGTENGKMTGYSSPEKAS 1229
1192 AGLEMQAHYDVFVWADNTTINDNYHTVTFDALGRVTSFRFWGTENGKQGYTPAENET 1251
1230 --FSPPSDVNAALIKPLPVAQCQVYAPESWMPVLVSKTFNRLAEQDWOKLYNARIITE 1287
1252 VPFVTPTTDDALAKPGIPVAGLMVYAPLSWM---VQASFSNDGEL-YGELKPAIGIITE 1307
1288 DGRICTLAYRRWQSQ--KAIPQLISLNNNGPRLPPLHSLTLTDRYDHPDPEQQRQVWF 1345
1308 DGYLLSLAFRRWHNNPAAAMPKOVNSQN-----PPHVLSTVITDRYDADPEQQRQVTF 1362
1346 SDGFRLLQAAARHEAGMARQRNEDGSLI 1374
1363 SDGFRITLTAVRHESGEAWRDEYGAIV 1391
```

RESULT 14

```
Q4ZP56 PSESY PRELIMINARY; PRT; 1447 AA.
AC Q4ZP56;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Salmonella virulence plasmid 65kda B protein.
GN ORENAMES=PeVr 4036;
OS Pseudomonas syringae pv. syringae B728a.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
OX NCBI_TaxID=205918;
(1)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RG DOE Joint Genome Institute;
RA Chain P., Larimer F., DiBartolo G., Copeland A., Lykidis A., Trong S.,
RA Nolan M., Goldsman E., Thiel J., Malfatti S., Lapidus A., Dettler J.C.,
RA Land M., Richardson P.M., Kyrpides N.C., Ivanova N.;
RT "Comparison of two complete genome sequences of Pseudomonas syringae
pv. syringae B728a and pv. tomato DC3000."
RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).
(2)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Loper J.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
(3)
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=B728a;
RA Feil H., Feil W.S., Lindow S.E.;
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
DR EMBL; CP000075; AAX39066.1; -; Genomic_DNA.
DR InterPro; IPR004048; Reg_chrom_condens.
DR InterPro; IPR003284; Sal_SpV8.
DR Pfam; PF03534; SpvB; 1.
DR PRINTS; PR01341; SALSPVBPROT.
DR PROSITE; PS00626; RCC1.2; UNKNOWN 1.
SQ SEQUENCE 1447 AA; 161914 MW; B76B60FB17F29009 CRC64;
```

Query Match 28.8%; Score 2277.5; DB 2; Length 1447;
Best Local Similarity 36.1%; Pred. No. 5.2e-136;
Matches 544; Conservative 225; Mismatches 595; Indels 143; Gaps 37;

SQ	SEQUENCE	1446 AA; 162070 MW; 8DBAAD65CDC8949 CRC64;
	Query Match	28.8%; Score 2274; DB 2; Length 1446;
	Best Local Similarity	35.7%; Pred. No. 8.7e-136;
	Matches 550; Conservative 220; Mismatches 582; Indels 188; Gaps 39;	
Qy	2	QNSQDFSTLSELPKGGGAIATGMEALPTPTGDCMAALSPLPISACRGVAPAPFTLVNS 61
Db	13	QSLQPAVATPL-LPKGGAIQISQKGGWGSVTSAAASLEVALPISPRGVAPALSLSYQS 71
Qy	62	GAGNSPFLGWDNCNMVTRRTHRGVPHYDFTDFLQPEGEVLVAVDQPRDE----- 113
Db	72	TSNGVFGGLWNLNTSKVARRASQVPTYANDDLIFGPGGDVCL---PERDSCALYSTQ 128
Qy	114	-STLQGINLGNFTVTGYRSRLESFHSRLEWQPKTKDFTWLIYSPDGOVHLLGSPQ 172
Db	129	VSRYNHDNLNATYQVRYFSRVEGAFLRIEHW-RVDIADPGFWLIHGADSLNLYGRRTS 187
Qy	173	ARISNPSTTQTAOWLLEASVSSRGEIYYOYRAEDDTGCEADBITHLQATAQRYLHV 232
Db	188	SRIADPADMNKVAEVLDESNNALGEHLIYKPEDHQGLAED---HPRNFRADQYLSRV 244
Qy	233	YYGNRTASETL-----PGLDSAPSQADWLFYLFYDYGERSNNLKTTPAFSTTGSWLCRQ 287
Db	245	RYGNAKAHYVLYLWQEDSLDGLL-----WHFDLIFDYQDRTRSEPPPEYDEQFTWVRS 299
Qy	288	DRFRYBYGFIIRYRCLRQVLMYHHLQALDSKITEHNGPTLVSRLLIYNDESIASTLV 347
Db	300	DPHSSFAFGFELGNLRCQVLMFHH---FPNELGE--APLLTRRLLEHLQTLTGYNLL 354
Qy	348	FVRVVGHEQ--DG---NVVTLPLELAYQDFSPRHHAHWPMDVLNFAINAIRQWLVDL 401
Db	355	---SAAHSQAWGDWTRVDOQPVPQYQYTDFS--LESGIYTPLEPMAGLNDGQYQLVDL 410
Qy	402	KGEGPLGLLYODKGAWYRSARLGEIGSDAVTWKMQPLSVIPSLOS----NASLVGIN 457
Db	411	YGDGLPLGLYRDDKAWLYREPVRDTTGGADAVAYGACQPLPRIPTADSATPVRTLDTLT 470
Qy	458	GDGQLDWITGPGLRGVHSQBPDSWTRPTPLNALPVEYTHPRAQLADLMGAGLSLVLI 517
Db	471	GDGRLDWVVAQPGWAGFTFLNPDRSWSNYATFSAPFAFFHPQOGMADLVGDGLSLDALI 530
Qy	518	GPKSVRLYANTR-DGFAKGKDVQSGDITLTPVPGADPKLVAFSDVLGSGOAHLVVEVSAT 576
Db	531	GPRSVRLYANRADAQFAAATDI PHDED-RLPLLSDSSTELVAFSDLLGTGQQLIRIRHN 589
Qy	577	KVTCWPNLGRGFGQP---ITLPGFSQPAFENPAQVYLADLDGSGPTDLIYVHTNRLDI 633
Db	590	EIRVWPNLGRGFGKGLFATLP-YTYEA--FDSSQVRLADLDGSGASDVLYLQADGFQV 646
Qy	634	FLNKSNGFAPPVTLRFPPEGLRFPDHTCOLQWADVQGLGVASLIISVPHMSPHHW----RC 689
Db	647	FMNRGGNGLAAPPDQRPWPEGVRYDFCQFSAVDLLGLGFSLSVLTVPHMAPLHWSLYAA 706
Qy	690	DLTN--MKPMLLNEMNMNMVHHTLYRSSQFWLDEKAAALTGQTPVCVLPPIHTLW 747
Db	707	DRAGSVKRPYLLKASDNNLGAAGEVSTRSSAQEWLDEKNELRAAGSVAVSELFPVHVVV 766
Qy	748	QTETEDISGNKLVTLTRYARGAWDGREREPFGYVEQTDSSHQAQGNAPERTPPALTK 807
Db	767	QRTMQDKVTGNTLQLFRYRQGFYDPREREPFGGLLLQTDTSQ--QBEDFTAPVLNK 825
Qy	808	NWYATG-----LPVIDNALSTEYWDQDQAQFAGFSRFTTWQ 843
Db	826	TWFTGTGYPARPCTDHRSDILLARLPGEHVLRLDAATQTQE----- 867
Qy	844	DNKDVPILTPEDDNGRYWFNKGLLRSELYGLDDSTNKHVPYTVTEFRSQVRLQHTD 903
Db	868	-----PITDADATLQEWARALSGSVLSEVFGLDASQRPVLYSTRSCRVLVRLQALS 922
Qy	904	SRYPLWSSV-----VESRNYHYE-RIASDPQCSQNIITLSSDRFQOPLKQLSVQYPRRQOP 958
Db	923	AHRE--YASMLPLSLEVITYRYEAEELEDPMCEHSLNLANWDYRGSTLHVSVNYARRKKP 980

Qy	959	A-INLYPDTLPLDKLANSYDDQORQLRLTYQSSWHHLTNNVTVRVLGLPDSTRSDIFTYG 1017
Db	981	GDAPPFADPHQQQWWEASHDDAQOQFYLNEMHABAIYLDSPQSWRLGLGIPYTRGDAMLIP 1040
Qy	1018	AENVPAGLNLLELLSDXNSLIADDPKREYLQOQKTAFT---DQNTTTPLOTFPTQALIAF 1074
Db	1041	ASALTPAQISVYEQPADPSGPPFA-TLPRTLTSLSVQRYIGCGDGE-----ASFQALADA 1092
Qy	1075	TETTVNQSTLSAPNGSIPSDKLSLTLEQAGYQOQNTNLYFPRTGEDKVVVAHHGYDYDTA 1134
Db	1093	VETAELDHALSAYERVMDSVTLAEKLVEIGYQOMPSPFLPADSLN-LWSVKRGFATVAGQ 1151
Qy	1135	AQFWRPQKQSNLTQTKITLIWDANYCVVQTRDAAGLTTSKAYDWRFLTPVQLTDINDN 1194
Db	1152	BHFFHPTQFPRTRSHGWSLVEYDAYHLFATRIITDPADCVTTAEYDVRVLQPKRIIDPNQN 1211
Qy	1195	QHLITLDALGRPITFLRFWGTENGKMTGYSSPEKASFSPSDVNAAIELKKPLPVAQCQVY 1254
Db	1212	ROEADYDAFGRVWATSVYGTGELGEAVGFPPLNRAGHYWASAGEVALQ----- 1258
Qy	1255	APESWMPVLQKTFNRLAEQDMQKLYNARIITEDGRICTLAYRRWVSQKAIPOILISLN 1314
Db	1259	-PE-----YALGRQ-----ASALYYDGN-----TVLGLVH 1282
Qy	1315	NGPRLPHSLTLTTDRYDHDPEQOIROQVVSFSDGFRLLQAAARHEAGMARQNRNEDGSLI 1374
Db	1283	-----IPLATAVLVADRYPEDLDKQIRISMASIDGRTLQTRQKVEDGDAYSVDENWGLE 1338
Qy	1375	I-----NVQHTENRWAVTGRTEYDNKGQPIRTYQPYFLNDWRYVYSNDSARQEKAYADTH 1429
Db	1339	LVDGKPKIVHASPRWRISERVEYNNKGLAVRVYRPFANSHLYYNDASIRSON--IVDKQ 1396
Qy	1430	VYDPIGREIKVITAKGWFRTLTFTPWETVNEDENDTAAEV 1469
Db	1397	FYDPLGRPTTITAKGWMRRQYRVWYTTISEDENDTAAEV 1436

Search completed: February 16, 2006, 21:42:55
Job time : 205.739 secs

Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	6031.5	76.3	1476	6	US-10-647-956A-4	Sequence 4, Appli
2	169.5	2.1	14130	7	US-11-175-689-9	Sequence 9, Appli
3	162.5	2.1	16990	7	US-11-175-689-7	Sequence 7, Appli
4	150	1.9	2228	7	US-11-124-367A-270	Sequence 270, App
5	150	1.9	3377	7	US-11-124-367A-271	Sequence 271, App
6	150	1.9	3841	7	US-11-124-367A-272	Sequence 272, App
7	139	1.8	2315	7	US-11-113-424-53	Sequence 53, Appli
8	136	1.7	2769	7	US-11-113-424-14	Sequence 14, Appli
9	131.5	1.7	2105	7	US-11-052-554A-173	Sequence 173, App
10	127.5	1.6	2204	7	US-11-052-554A-174	Sequence 174, App
11	127	1.6	2591	6	US-10-453-372-718	Sequence 718, App
12	127	1.6	2602	6	US-10-453-372-716	Sequence 716, App
13	127	1.6	2617	6	US-10-453-372-666	Sequence 666, App
14	127	1.6	2617	6	US-10-453-372-732	Sequence 732, App
15	127	1.6	2617	6	US-10-453-372-734	Sequence 734, App
16	127	1.6	2617	6	US-10-453-372-736	Sequence 736, App
17	127	1.6	2617	6	US-10-453-372-738	Sequence 738, App
18	127	1.6	2617	6	US-10-453-372-740	Sequence 740, App
19	127	1.6	2617	6	US-10-453-372-742	Sequence 742, App
20	127	1.6	2617	6	US-10-453-372-744	Sequence 744, App
21	127	1.6	2617	6	US-10-453-372-746	Sequence 746, App
22	127	1.6	2617	6	US-10-453-372-748	Sequence 748, App
23	127	1.6	2617	6	US-10-453-372-750	Sequence 750, App
24	125.5	1.6	1579	7	US-11-052-554A-9	Sequence 9, Appli
25	125	1.6	1323	6	US-10-517-939-312	Sequence 312, App

Db 5510 L-----FSVMVSLAALWRW--AGITPDAVIG-----HSQ 5536
Qy 426 GEIGSDAVTEKMOB-LGVIPSLQSNASLVINGDQOL-----DWIT 467
Db 5537 GEIAAAHVAGALTPEAAAVVALSRV-LTDLAGAGAWASVLSPEEPTQLLARWDGKIT 5595
Qy 468 GPLRGYHSQRDPDSWTRFTPL-----NALPVEY-----THPRAQLAD----- 505
Db 5596 VAAVNGPASAVVSGDTTATITELLITCEHENIDARAIPVDYPSHSPYMEHIRHQFLDELPE 5655
Qy 506 -----LMAGLSDLAVL 516
Db 5656 LTPRPSTIAMYSTVDGEPHDTAYDTTWTADYWRNIRNTRFHTDVAALIGAGEQVPLE 5715
Qy 517 IGPKSVRLYANTRDGFAGKDVQVQ-----SGDITLVPVGPADPRKLVAFSVDVLSGQAHLEVEV 573
Db 5716 LSPHPVLTOALT-----DTVEQAGGGAAPALRKDRPDVAFAAALQOLHCHGISP 5767
Qy 574 SATKVTCPNIGRGRFGQPIITLPGFS-----QPATBNPAQVYLAD-----LQSGGPTD 622
Db 5768 SNNVLYCQ-----ARPLTLPTVAFQHQRYWLLPTAGDFSGANTHAMHPLLDTA--TE 5817
Qy 623 LI-----YHTNRL-----DIFLANK-----SGNGFAEPTVTLRPPGELRFD-----H 658
Db 5818 LAENRGWFTGRISPRTPQWLNHEHAVESAVLPFGTGYE-LALHVADRAGYSVNNELIVH 5876
Qy 659 T-----COLQADVOGLGVASLILSVPMSPH--HWRCDLTNMKP-WLLNEMNN 704
Db 5877 TPLLAGHDTADLQITVTDMDGRQSLNI-----HSRPHIGHDNVTYTGDEQEFEWLH----- 5929
Qy 705 NMGVHTLRLYRSSQFWLDEKAAALTTCQT-----PVCYLPFPPIHTLMOTETED----- 753
Db 5930 -----ASAVLTAQTDDHNLPLTPVWPBPPTAGAEVDDFYDDL 5968
Qy 754 EISGNKLVTLRYARGAWD-----REREPFGY----- 783
Db 5969 AAQYNTYGTPTQGVQRIWRDHATPDVIAVEVELPEDTIDIGYTHPALFDAAHLPLAL 6028
Qy 784 ---VEQTDHQALQAGNAPERTPAL-----KNWYATGL 814
Db 6029 QPPTNDTDDTWTADTGDOVRLPYAFTGISLHATHATRLVRLTRTGADAITVHSDTTGA 6088
Qy 815 PV--IDNALSTEYWRDQAFAGSPRFT-----TWQDNKDVLPTPEDDSRYWFNRL 865
Db 6089 FVAIIDSLLIT--RPLTTATGSAATTAAAGLLHLSWPPHPTDITDITDITD-----AL 6138
Qy 866 KGOLLRSE-----LYGLDSTNKHVPYTFEFSQVRRLQHTDSRYPLVSSVWESRN 918
Db 6139 RYQVIAEPTQOLPRYLHDLHTSTDLHTTTEADV-----VWMPVPVPSNE 6183
Qy 919 YHYERIASDPOCS-----QNITLSSD-----RFG-----QPLKQL 948
Db 6184 ELQAHQASDVAVSSRIHTLRTQTLIVQDNLTHPTTGTTRLVIVTRHGVSATSDPVDL 6243
Qy 949 S-----VOYPRRQQA-----INLYPDTLPDKLLANSYDDQOQLRLTYQQSSWH 993
Db 6244 AHAAMGLIRSAQNEHGRFTLLDDTNDNSDTLTALTLPRENO-----LAIRRDIT 6298
Qy 994 -----HUTNNTVRVLGPDSTRDIFTYG-----AEN-VPAGGLNELLSKNSLI 1038
Db 6299 IPRLRHSSDGLALTAPVVVDPEGTGLITGGTGLGALFAHLVSAHVRHLLTSRR----- 6355
Qy 1039 ADDKREYLGQKATYD-GQNTTLPQTPTRQALIAFTETTVENGSTLSAFNGSIPSDKL 1097
Db 6356 -----GQAHGATDLOQLRDL-----GAHTITACDISDEALALVNSVPTQHR 6401
Qy 1098 STTL-----EQAGYQQTNYLFPRTGEDKVVVAHHGVYTDYGTAA--QFWRPQK 1142
Db 6402 LTAVVHTAAVLADTPVTELTGDLQDLVAPKI--DAAMQLQLTYEHLNSAFIMF----- 6454
Qy 1143 QNSTQLTKITLIWANDYCVVQVTRDAAGLTTSAKYDWRFLTTPVQLTDINDNQHLITLDA 1202
Db 6455 ---SSMAGWISPGQGNYA-----AANTALDALADYR-----HR 6485

Qy 1203 LGRPITLRFWG-----TENGMKTYGSSPEKASFPSPSDVNAALIEL 1242
Db 6486 LGLPATSLANGWQTHHTGLTAHLTDVLDLARMTRGLMPLATSHGLALFDAALATQOPVSI 6545
Qy 1243 KKPLPVAQCQVYAPESWM-PVLS-----OKTFNRLAEQDWOKLYNARI 1284
Db 6546 PAPINTHTLARHARDNTLAPILSALITTPRRRAASAAATDLAARLNGLSPOOQOQTLATLV 6605
Qy 1285 ITEDGR1-----CTLAYRRWQSKAIQOLISLNN-GPRLPPHSLTLTTDRYD 1332
Db 6606 AAATATVLGHTPESIPATAPKDLGIDSLTALERNTLTHTGDLLEP-----TLIFD 6659
Qy 1333 HDPEQIQIRQVVS-DGFRLLQAAARHEAGMARQEDGSLII-----NQVHTNR 1383
Db 6660 HPTPHAVAEHLLEQIPGIGALVPAPVIAAG-----RTEEPVAVVMACRFPFGVASADQL 6715
Qy 1384 W--AVTGRTEDN-----KGQPI-----RTYQPY--FLND 1409
Db 6716 WDLVIAGRDVVGNFPADRGMDVEGLFDPDPAVGKTYTRYGAFLOD 6761

RESULT 3

US-11-175-689-7

; Sequence 7, Application US/11175689

; Publication No. US20060024806A1

; GENERAL INFORMATION:

; APPLICANT: STINEAR, TIMOTHY P.

; APPLICANT: COLE, STEWART T.

; APPLICANT: LEADLAY, PETER F.

; APPLICANT: SMALL, PAMELA L.C.

; APPLICANT: JOHNSON, PAUL D.R.

; APPLICANT: JENKIN, GRANT A.

; APPLICANT: DAVIES, JOHN K.

; APPLICANT: HAYDOCK, STEPHEN F.

; TITLE OF INVENTION: THE MYCOLACTONE LOCUS: AN ASSEMBLY LINE FOR PRODUCING

; FILE OF INVENTION: NOVEL POLYKETIDES, THERAPEUTIC AND PROPHYLACTIC USES

; FILE REFERENCE: 03495.0329-01

; CURRENT APPLICATION NUMBER: US/11/175,689

; CURRENT FILING DATE: 2005-07-07

; PRIOR FILING DATE: 2004-11-15

; PRIOR APPLICATION NUMBER: 10/987,592

; PRIOR FILING DATE: 2003-11-14

; NUMBER OF SEQ ID NOS: 36

; SOFTWARE: PatentIn Ver. 3.3

; SEQ ID NO 7

; LENGTH: 16990

; TYPE: PRT

; ORGANISM: Mycobacterium ulcerans

US-11-175-689-7

Query Match 2.1%; Score 162.5; DB 7; Length 16990;

Best Local Similarity 18.5%; Pred. No. 0.034;

Matches 342; Conservative 186; Mismatches 631; Indels 685; Gaps 86;

Qy 4 SQDPSITSLUPKGGGAIITGMEALTPTG-----PDGMAA-----LSLPLPISAGRYA-- 52

Db 11465 NQDQASGLTAPHGPSQORVINQALANAGLTHDQDVAEANGTGTTLGDPPEASALHATY 11524

Qy 53 -----PAPTLNYSAGNSPFLGWDGNVMTIRRTTHFGVP---HYDETDFLG-PE 100

Db 11525 GHHTTPOPLWLGSIKSNIGHTQAAAGAAVGVVMIQAITATLATLHVDOPSHIDWSS 11584

Qy 101 GEVLVV-----ADOPRDESTLOGINLGATFTVTGYRSLESHFSLEYWPKTKTKT 152

Db 11585 GTVRLLEPTIOWPNTDHPRTAAV-----SSFGISGNTAHLILO-----QPPTDPT 11630

Qy 153 DFWLIYSPD---GQVHLGLKSPQARISNPSTOTTAQWLLLEA---SVSSRGEQIYYQYR 205

Db 11631 Q-----TPNTTGTGSDPAVGPVGV-----LWVPLSARSPAGLSAQAARLYQHLS 11676

Qy 206 AEDDTGCEADEITHHLOAQAORYLH-----IVVY--GNRTASETLPLGDLGSPASQADM 256

Db 11677 AHPD---LDPIDVAHSLATRRSHPHRATITTSIEHSENNHDTTDAALHALANNGTHP 11734
 Qy 257 LFYLVDFYDGERSNLKTTP-----AFSTTGWMLCRQDRSRVEY---GPEIRTRLCR 306
 Db 11735 LL-----SRGLLTPQSGKTVFVPPGQS-----QYPMGAD-----LYR 11769
 Qy 307 QVLMY-HHLQALDSKITEHNGPTLVSLRLILNYDESATLTVFVRVYRGHEODGNVTVLPP 365
 Db 11770 QFPFAHALDACDALQFTGWSVLAVL---HDEPEAPS-----LERVD-----VQPV 11815
 Qy 366 LELAYQDFSRRHAHQPMVDLANFNALQRQLVDLKGEGPLGLLYQDKGAWYRSQRL 425
 Db 11816 L-----FSWVSLAALRW--AGITPDAVIG-----HSQ 11842
 Qy 426 GEIGSDAVTWKMP--LSVIPSLQSNASLVINDGGQL-----DWTIT 467
 Db 11843 GEIAAAHVAGALTLPAAAVVALRSRV-LITLAGAGAMASVLSPEEPLTQLLARWDGKIT 11901
 Qy 468 GPGLRGYHSQRPDGSWTRFTPL-----NALPVEY-----THPRAQLAD--- 505
 Db 11902 VAAVNGPASAVVSGDITAITELLITCBHENIDARAIPVDPSHSPYMEHIRHQFLDELPE 11961
 Qy 506 -----LMGAGLSDLVL 516
 Db 11962 LTPRPSTIAMYSTVDGPHDTHAYDTTWTADYWNINRTVRFHTVTAALLGAGEQVPLE 12021
 Qy 517 IGPKSVRLYANTRDGFAGKDVQVQ---SGDITLVPVGDPRKIVAFSDVLGSGQAHUVEV 573
 Db 12022 LSPHPVLTAIT-----DTVEQAGGGGAAPALRRKRPDAVAFAAALQLHCHGISP 12073
 Qy 574 SATKVTCPNLGRGFGPITLPFGS-----QPATENPAQVVLAD---LDGSGPTD 622
 Db 12074 SWNVLYCO-----ARPLFTYAFQHORYWLLPTAGDFSGANTHAMHPLDPA--TE 12123
 Qy 623 LI-----VYHTNRL-----DIFLNC-----SGNGFAEPVTLRFPEGLRFDHTCQLQMA 665
 Db 12124 LAENRGWFTGRIISPRTPQWLNHAEVSAVLPFGTGFVE-----LALHVA 12168
 Qy 666 DVQGL-GVASILSVPHMSPHHWRCDL-----TNMKPMLNEMNNMNGVHHHTLRYRSS 717
 Db 12169 DRAGYSSVNELIVHTPILLAGHDADLIQITVTDGDMGROSLNTHSPHIGHDNTTTCDE 12228
 Qy 718 SQFWLDEKAAALTTGQTPVCVLPPIHTLMQTEFDEISGNKLTTLRYARGANDGRE 777
 Db 12229 QPEWVLHASAVLTAQTTHDNLPL----- 12252
 Qy 778 FRGFGYVEQTDHQLAQGNAPERTPPALTKNMYATGLPVIDNALSTBYRDDQAFGS- 836
 Db 12253 -----TPVPWPPCTA-----ALEVDDFYDDLAAQGVNY 12281
 Qy 837 -PRF-----TTWQDN-----KDVPLTPEDDNRWYFNRAKLGQLLSLYGLDDSTNK- 883
 Db 12282 GPTFGQVIRWRDHATPDVIAEVELPDTDIDGYIHPALFDAALHPALLALTPPTNDT 12341
 Qy 884 -----HVPYVTTEP-----RSQVRLO-----HTD--SRYPVLNWSV 913
 Db 12342 DDTWTADGQVRLPYATGILHATHATRLVRVLTGTGADAITVHTSDTTGAPV---AI 12398
 Qy 914 VESRNYHYRIASDPQS-----QNITLSSDRFQOPLKQLSVQY 952
 Db 12399 IDS-----LITPLITATGSAPATTAAGLLHLSWPPHPTTDTTDDTDLARYQVIAE 12451
 Qy 953 PRROOPA-----INLYPDTLP--DKLLANSYDDQQLRLTYQQSSWHHLTN 997
 Db 12452 PTQQLPYRLDLHTSTTEADVVVWVPVPSNEELQAHOASD-----TAVSSRIHTLTR 12504
 Qy 998 NTVEV-----LGLPDS-----TRSDIFTYGAENVP-----AGGNLELLSPKNSLIA 1039
 Db 12505 QTLTVQDQWLTHTPDTTGTRLVIVTRHGVSSTAHPDVPDLAAHAWGL---IRSAQN--- 12557
 Qy 1040 DDKPREVLGQOKYATDQGN-----TTPLOTPTTRQALIAFTETTVF-----NOSTLSAFN 1089

Db 12558 -----EHPGRFTLLDNTNSDTLTATTALTPTRENQAIARRDTIHIPRLTRHSSDGALT 12612
 Qy 1090 GSIPSDKSLTLEAGYQQOYNILFPRTGEDKVVAAH-----HGYTDYG----- 1132
 Db 12613 APVVVDEPGVTLTGTGTGLGALFAEHLVSAHGVRLHLLTSRRGPOAHGATDLOOQLTDL 12672
 Qy 1133 -----TAAQFWRPQKOSN-----TQUTG-----K 1151
 Db 12673 GAHVTITACDISPEALAAALVNSVPTQHRILTAVVHTAAVLADTPVTLTGDLQDOVLAPK 12732
 Qy 1152 ITLIW-----DANYCVVQTRDAAGLTTSKAYDMRFLFPVQLTIDNNOHLITLTDALG 1204
 Db 12733 IDAAWQLHQLTYEHLNSAFIMFSGMAGWISPGCGNYAAANTALDALADYRH-----RLG 12787
 Qy 1205 RPTILRFGW-----TENGMTCYSSPEKASFSPSPDVNAAIELKK 1244
 Db 12788 LPATSLANGWYQHTGTHTAHLTDVLDLARMTRLGLMPTATSHGLALFDAALATGQPVSIIPA 12847
 Qy 1245 PLPVAQCQVYAPESWM--PVLS-----OQTFNRLAEQDMQKLYNARIIT 1286
 Db 12848 PINTHTLARHARDNTLAPILSALITTPRRRAASAATDLAARLNGLSPOQOQOTLATLVAA 12907
 Qy 1287 EDGRI-----CTLAYRRWVQSKAIPOLISILNN--GPRLPPHSLTLTTDRYDHD 1334
 Db 12908 ATATVLGHHTPESISPATAFKDLGIDSLTALRLNTLTHNTGLDLP-----TLIFDHP 12961
 Qy 1335 PEQIIRQOVVPS--DGFRLILQAAARHEAGMARQNEGSLII-----NVQHTENRW- 1384
 Db 12962 TPHAVAEHLLEQIPGICALVPAPVIAAG---RTEEPVAVVGMACRFPGGVASADQLWD 13017
 Qy 1385 -AVTGRTEYDN---KGQPI-----RTYQPY--FLND 1409
 Db 13018 LVIAGRDVVGNFADRGWDVEGLFDPDPDAVGKTYTYGAFLLD 13061

RESULT 4

US-11-124-367A-270
 ; Sequence 270, Application US/11124367A
 ; Publication No. US20060024700A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Michele Cargill
 ; APPLICANT: Hongjin Huang
 ; TITLE OF INVENTION: Genetic Polymorphisms Associated with
 ; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
 ; FILE REFERENCE: CL001519. ORD
 ; CURRENT APPLICATION NUMBER: US/11/124,367A
 ; CURRENT FILING DATE: 2005-05-09
 ; PRIOR APPLICATION NUMBER: US 60/568,846
 ; PRIOR FILING DATE: 2004-05-07
 ; PRIOR APPLICATION NUMBER: US 60/582,609
 ; PRIOR FILING DATE: 2004-06-25
 ; PRIOR APPLICATION NUMBER: US 60/599,554
 ; PRIOR FILING DATE: 2004-08-09
 ; NUMBER OF SEQ ID NOS: 34460
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 270
 ; LENGTH: 2228
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-11-124-367A-270

Query Match 1.9%; Score 150; DB 7; Length 2228;
 Best Local Similarity 17.5%; Pred. No. 0.013;
 Matches 261; Conservative 199; Mismatches 498; Indels 530; Gaps 67;
 Qy 220 HLOATAQRYLHVIVYGNRTASETLPGLDGSAPOADWLFYLVDFYDGRSNLKTTPAFST 279
 Db 366 HLKFTQDVSAVV--KENSTEATLAVITACNPINEPLFYHILNPDRRFKISRTSGVLST 424
 Qy 280 TGSWLCHQDRFSYEYGFETRLCRQVLMYHHLQALD--SKITEHNGPTLVSRLLNY 337
 Db 425 TGTPDFREQ-----EAFDVVVEVTEEHKPSAVARVV--- 456

Qy	338	DESALASTLVFRRVGHBDQGV---VTLPPELELAYQD-----	372
Db	457	-----VKVIVEDQNDNAPVFNLPYYAVVKVDTEVGHVIRYVTVADRDGRNGE	505
Qy	373	---FSPRRHAWQPMVDLANFNAIQRLQVLDLKGGLPGLLAYQDKGAWYRSQRLGEIG	429
Db	506	VHYTLKEHHEHQ-IGPLGEISLKKQFELDTLNKEYLVTVVAKOG-----	550
Qy	430	SDAVTWKMQPLSVI-----PSLQSNASLVDINGDGLDWMVITGPCLR	472
Db	551	NPAPSAEIVIPITVMNKAMPVEKFPFYSAETASIQVHSPVVHQAN-----SPGLK	603
Qy	473	GYHSQRDPGSWTRT-----PLNAL-PVEY-THPPRAQLA-----DLMGA-----	509
Db	604	VFYISITDGPFSOFTINTGVINVIAPLDPEAHAPYKLSIRATDSLTGAHAEVFVDIIV	663
Qy	510	-----GLSDPLVLIGPKSVRLYANTRD-----	530
Db	664	DDINDNPPVFAQQSYAVTLSASVIGTSVGVVRATDSDSEPNRGISYQMGFNHKSHDHP	723
Qy	531	-----GPAKGKDVVQSGDITLPPVGADPRKLVAFSDVLGSGOAHLEVSATKVTWCW	581
Db	724	HVDSSITGLISLRLTDYEQSRQHTIFVRANDGGMPTLSSDV-----IVTVDTDLNDN	776
Qy	582	PNLGRGRFGQDITLPGFSQPATE---FNPAQVYIAD-----LDG-----	617
Db	777	PPL----FEQIYEARISEHAPGHGFVTCRAYDADSSDIDKLOYLSILSGNDHKHFVIDS	832
Qy	618	SGPTDLIYVHTNPLDIFLNKSGNGFAEPVTLRPEGLRDHDTCOLQWADVQGLGVASLI	676
Db	833	ATGIITLSNLHRHALKFPYS-----LNLVSVDGV-FRSSTQVHVTVTGG-----N	876
Qy	677	LSVPHMSGPHWRCDLTNNKP---WLLNEMNNNG-----VHHTLYRSSSQFWLDEKA	726
Db	877	LHSPAFLONEYEVELAENAPLHTLVMEKTIJDGSGIYGHVYHI VNDPAKDRFYINERG	936
Qy	727	AALT-----GQPPVCYLPFPIHTLMOTETEDBISGNKLVT 763	
Db	937	QIFLEKLDRETPAEKVISVRLMAKDAGGKVAFC-----TVNVILTDDNDNAPQPRAT	989
Qy	764	LY-----ARG-----AWDRERERFGYGVQTDSDHQLAOGNAPERTPALTKN	808
Db	990	KYEVNIGSSAAKGTSVVKVLASDADGGSNADITYAHEADSESVKENLEINKLSGVITTK	1048
Qy	809	WYATGLP-----VIDNA-----LSTEYWRDQAFAGFSPRFTTQONKDQVPL	850
Db	1049	ESLIGLENEFTFFVRADVNGSPSKESVVLVYVKILPEWQLPKFSEPFYFTYTSEDVPI	1108
Qy	851	TPEDNSRYWFNRAKQQLRLSELYGLDDSTNKHVPYTVTFERSQV---RRLQHTDSRYP	907
Db	1109	GTEIDLIR----AEHSGTVLYSLVKGNTPESNRDES FVIDRQSGRLKLEKSLDHETTKW-	1163
Qy	908	VLWSSVVESRNYHYERTASDPOCSQNTLSSDRGQPLKQLSVQYPRRQOPAINLYP---	964
Db	1164	-----YQFSILARCTQDDHEMVASVD-----VSIQVXKANDNSGVFESSPYEA	1206
Qy	965	---DTLPDKLLANSYDQORQLRUTYQQSSWHLLTNNTVRLGLPDSRSDIFTYGAENV	1021
Db	1207	FIVENLFG-----GSRVQIRASDADSG-----TNGQV-MYSLDQSQSVIEFSPAINM	1254
Qy	1022	PAGGINL---ELLSDKN-----SLIADKPREYLQOQKATYDQGNTTLPLOTPTQALIAF	1074
Db	1255	ETGWITTLKELDHKRDNYQIKVVASDH-GEKIQLSSTAI VDVTVTVDNDSPPR-----	1307
Qy	1075	TETTVFNQSTLSAFNGSIPDK-----LSTT---LEQAGYQOTNYLFPRTGEDKVVW	1123
Db	1308	-----FTAEIYKGTVEDDPQGVIAILSTTADSEINRQVTFPI---TGSDPL--	1354
Qy	1124	AHGYDYDGTAAQFWR-----POKOSNTQLTGKTLIWDANYCVVQTRDAAGLTTS	1175
Db	1355	---GQFVETIQNEKVVVKPLDRKRDNYLLT---IT-----ATD	1390
Qy	1176	AKYDWRELTVPQLDTINDNOHL-----ITLDALGRPTLTLRFWGTGTE-----NGKMTG	1221

Db	1391	GTSSKAIVVEVKYLDANDNSPVCEKTYLSDTIPEDVLPGLKIMQISATDADIRSAEIT-	1449
Qy	1222	Y-----SPEKASFPSSDVNAAIEKKPLPVAQCQVYAPESWMPVLSQKTFNRLAEQDQW	1277
Db	1450	YTLGSGAEKFLNPDTG-----ELKTSIPL-----DREQ	1480
Qy	1278	KLYNARIITED--GRICTLAYRRWVQSQAIPOLISLNL-NGPRL--PPHSL-----	1324
Db	1481	AVYHLLVRATDGGGRFC-----QASIVLTLEDVNDNAPEFSADPVAITVFENTEP	1530
Qy	1325	TUTTDYRDHPQOIRQWVFS-----DG-----FGRLLQNA-----A	1357
Db	1531	GTLTRVQATDAGLNRLKILYSLIDSADQFSINELSGIIOLEKPLDRELQAVYTLSLK	1590
Qy	1358	RHEAGMARORNEGSLIINVOHTENRWAVTGRTEYDNKQPIRTYQPY	1405
Db	1591	AVDQGLPRRLTATGTIVSVLDI-----NDNPPVPEYREY	1625
RESULT 5			
US-11-124-367A-271			
Sequence 271, Application US/11124367A			
Publication No. US20060024700A1			
GENERAL INFORMATION:			
APPLICANT: Michele Cargill			
APPLICANT: Hongjin Huang			
TITLE OF INVENTION: Genetic Polymorphisms Associated with			
TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof			
FILE REFERENCE: CLO01519.ORD			
CURRENT APPLICATION NUMBER: US/11/124,367A			
CURRENT FILING DATE: 2005-05-09			
PRIOR APPLICATION NUMBER: US 60/568,846			
PRIOR FILING DATE: 2004-05-07			
PRIOR APPLICATION NUMBER: US 60/582,609			
PRIOR FILING DATE: 2004-06-25			
PRIOR APPLICATION NUMBER: US 60/599,554			
PRIOR FILING DATE: 2004-08-09			
NUMBER OF SEQ ID NOS: 34460			
SOFTWARE: FastSeq for Windows Version 4.0			
SEQ ID NO 271			
LENGTH: 3377			
TYPE: PRT			
ORGANISM: Homo sapiens			
US-11-124-367A-271			
Query Match 1.9%; Score 150; DB 7; Length 3377;			
Best Local Similarity 17.5%; Pred. No. 0.025;			
Matches 261; Conservative 199; Mismatches 498; Indels 530; Gaps 67;			
Qy	220	HLQATQRVLHIVYGNRTASETLPLGDSAPSQADWLFLVFDYGERSNLTKTPPAFST	279
Db	1976	HLKFTQDVTSNV-KENSTEAETLAVITAGNPINELPHYINPDRPKISRTSGVLST	2034
Qy	280	TGSQLCRQDRFSRYEGFEIRTRRLCRQLVMYHHQLALD--SKITEHNGPTLVSRILNY	337
Db	2035	TGTFPDEQQ-----EAFDVVVEVTEHKPSAVAHVV---	2066
Qy	338	DESAIASTLVFVRVGHGQGNV---VTLPPELAYQD-----	372
Db	2067	-----VKVIVEDQNDNAPEVFNLPFYAVVKVDTEVGHVIRVYTAVDSDSRNGE	2115
Qy	373	---FSPRHHAHQPMVDLANFNAIQRWLVDLKGELPGLLVQDKGAWVRSQRGLGEIG	429
Db	2116	VHYLYLKEHHHFQ-IGPLGEISLKKQFELDTLNKEYLVTVVAKDG-----G	2160
Qy	430	SDAVTWEKKOPLSVI-----PSLQSNASLVNDINGDQLDWITGPGLR	472
Db	2161	NPAFSAEIVPIVTKNAKMPVEKPFYSAIEASIQVHSPVVHVQAN-----SPEGLK	2213
Qy	473	GYSORPDGDSWTRFT-----PLNAL-PVEY-THPRAQLA-----DLMGA-----	509
Db	2214	VFSYITDGPSPSOFTTNGNTGVINVTAPLDFEAPKAYKLSIRATDSLTGHAHEVFVDIIV	2273

```
Qy 510 -----GLSDLVLIQPKSVRLVYANTRD----- 530
Db 2274 DDINDNPPVFAQQSYAVTLSEASVIGTSVQVVRATDSSEPNRGISYQMFNGHSHKSHDHF 2333
Qy 531 -----GPAKGKDVVQSGDITLPVPGADPRKLVAFSDVLSGQAHVLSVATKVTWCW 581
Db 2334 HVDSSSTGLISLLRTLDEYQSRQHTIFVRAVDGGMPTLSSDV-----IVTVDTDLNDN 2386
Qy 582 PNLGRGFGQPIITLPGFSQAPATE---FNPAQVYLAD-----LDG----- 617
Db 2387 PPL-----FEQIYEAREISEHAPGHGFTVCYKAYDADSSDDKLOYSTLSGNDHKHFVIDS 2442
Qy 618 -SGPTDLIYVHTNRDLDFLNKSGNGPAPVTLRPEGLRFDHTCQLOMADVQGLGVASLI 676
Db 2443 ATGIITISLNRHRAKPFYS-----LNLVSVDGV-FRASSTQVHVTVIGG-----N 2486
Qy 677 LSVPHMSPHWRCDLTNNKP---WLLNEMNNMG-----VHHTLYRSSSQFWLDEKA 726
Db 2487 LHSFAPLQNEVEVELAENAPLHVLMEVKTTDGDGSIYGHVYHI VDNFAKDRFYINERG 2546
Qy 727 AALTT-----GQTPCYLPFPPIHTLWQTEDEISGNKLVTT 763
Db 2547 QIFTEKLDRETAEKVISVRLMAKAGKVAFC-----TVNVILTDNDNAPQPRAT 2599
Qy 764 LRY-----ARG-----AWDGRERPRGFGYVEQTSQHLAQGNAPERTPPALTKN 808
Db 2600 -KYEIVTGSSAAKGTSVVKVLASDADSGNADITAIADSESVKENLEINKULSGVITTK 2658
Qy 809 WYATGLP-----VIDNA-----LSTEYWRDDQAFAGFSFRFTTWQDNKDVP 850
Db 2659 ESLIGLENEFFTFVRAVDNGSPSKESVVLVYVVKILPEMQLPKFSEPFYFTFTVSEDVPI 2718
Qy 851 TPEDNGRYFNWRAKQGLRSELYGLDDSTNKHKVPVTVTEFRSQV---RLQHTDRYP 907
Db 2719 GTEIDLIR-----AEHSGTVLSYLVKGNTPESNRDESFIIDQSGRLKLEKSLDHETTKW- 2773
Qy 908 VLWSSVESRNYHYERTASDPQCSQNTILASDRFGQPLKQLSVQYPRRQQAINALXP--- 964
Db 2774 -----YQFSILARCTQDDHEMWASVD-----VSIQVKAANDNPFVSESSPYEA 2816
Qy 965 ---DTPDKLIANSYDDQORLTYQOSSMHHLTNNTVRLGLPDSRSDIFTFYGAENV 1021
Db 2817 FIVENLFG-----GSRVIQIRASDADSG---TNGQV-MYSLDQSQSVIEBFAIMN 2864
Qy 1022 PAGGILN---ELLSKN-----SLIADKPRYLQOQKATYTDGQNTTPTQPTQALIAF 1074
Db 2865 ETGWITTLKELDHKDRNYQIKVVASDH-GEKIQLSSTAIYDVTVTVTVNDSPPR----- 2917
Qy 1075 TETTVFNQSTLSAFNGSIPSDK-----LSTT---LEQAGYQOQTNVLPRTGEDKVWV 1123
Db 2918 -----FTABIYKGTVSEDDPOGGVAILSTTDADSEENRQVTFYI---TGQDPL-- 2964
Qy 1124 AHGYTDYGTAAQFWR-----POKOSNTQLTKITLIWDANYCVVQVOTRDAAGLTTS 1175
Db 2965 ---GQPAVETIQNEKVKYVKKPLDREKDNVLLT---IT-----ATD 3000
Qy 1176 AKYDWRFLTPVLQFDINDNOHL-----ITLDALGRPITLRFWQTE-----NGKWTG 1221
Db 3001 GTFSSKAIVEVKVLADNDNSPVCEKLTYSDTIPEDVLPGLKIMQISATDADIRSNABIT- 3059
Qy 1222 Y-----SSPEKASFPSPDNNAAIELKPLPVAQCVQVAPESWMPVLQKTFNRLAEQDWQ 1277
Db 3060 YTLGSGAEKPKLNPDGTG-----ELKSTSTPL-----DREEQ 3090
Qy 1278 KLYNARIITED--GRICTLAYRRVWQSKAIPOLISLIN-NGPRL---PPHSL----- 1324
Db 3091 AVYHLLVRATDGGGRFC-----QASIVLTLEDVNDNAPFSADPAIVTFENTEP 3140
Qy 1325 -TLFTDRYDHDPQOQIRQVVFS-----DG-----FGRLLQAA-----A 1357
Db 3141 GTLLTRVQATDADAGLNKILYSLIDSADGQFSINELSGIITQLEKPLDLRELQAVYTLSLK 3200
```

```
Qy 1358 RHEAGMARQORNBGSLIINQVHTENRWAVTGRTEYDNKGQPIRTYQPY 1405
Db 3201 AVDQGLPRRLTATGTIVSVLDI-----NDNPPVEFYREY 3235

RESULT 6
US-11-124-367A-272
; Sequence 272, Application US/11124367A
; Publication No. US20060024700A1
; GENERAL INFORMATION:
; APPLICANT: Michele Cargill
; APPLICANT: Hongjin Huang
; TITLE OF INVENTION: Genetic Polymorphisms Associated with
; TITLE OF INVENTION: Fibrosis Methods of Detection and Uses Thereof
; FILE REFERENCE: CL001519.ORD
; CURRENT APPLICATION NUMBER: US/11/124,367A
; CURRENT FILING DATE: 2005-05-09
; PRIOR APPLICATION NUMBER: US 60/568,846
; PRIOR FILING DATE: 2004-05-07
; PRIOR APPLICATION NUMBER: US 60/582,609
; PRIOR FILING DATE: 2004-06-25
; PRIOR APPLICATION NUMBER: US 60/599,554
; PRIOR FILING DATE: 2004-08-09
; NUMBER OF SEQ ID NOS: 34460
; SOFTWARE: Fast-SEQ for Windows Version 4.0
; SEQ ID NO 272
; LENGTH: 3841
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-11-124-367A-272

Query Match 1.9%; Score 150; DB 7; Length 3841;
Best Local Similarity 17.5%; Pred. No. 0.03;
Matches 261; Conservative 199; Mismatches 498; Indels 530; Gaps 67;

Qy 220 HLOATAQRYLHIVYGNRTASETLPGDGSAPSQADMLFYLVDFDYGRSNNLTKPPAFST 279
Db 1979 HLKFTQDVYSAVV-KENSTEATLAVITAGNIPNEPLFYHILNPDRRFKISRTSGVLST 2037
Qy 280 TGSWLCHQDRFSRYEYGFETRRLCHQVLMYHQLQALD--SKITEHNGPTLVLSRLILNY 337
Db 2038 TGTPFDREQQ-----EAFDVVVEVTEEHKPSAVAHVV--- 2069
Qy 338 DESAIASTLVFVRVHGHEQDGNV---VTLPLELAYQD----- 372
Db 2070 -----VKIVEDQNDNAPFVNLPIYAAVVKYDTEGVHIVRYVTAVDGRNGE 2118
Qy 373 ---FSPRRHAWPMDVLANFNAIQRMQLVDLKGEGPLGLLYQDKGAWWYSRAQRLGEIG 429
Db 2119 VHYLYLKEHHEHFQ-IGPLGEISLKKQFELDTLNKEYLTVTVVAKDG-----G 2163
Qy 430 SDAVTWKMQPLSVI-----PSLQSNASLVNDINGQDQLDWITGQLR 472
Db 2164 NPAFSAEVIPIITVMNKAMPVPEKPFYSABIAESIQVHSPVHVQAN-----SPEGLK 2216
Qy 473 GYHSQRPDGSWTRFT-----PLNAL-PVEY-THPRAQLA-----DLMGA----- 509
Db 2217 VFISITDGPSPQTFINFTNCGVINVIAPLFEAPKLSIRATDSLITGAHAEVFVDIIV 2276
Qy 510 -----GLSDLVLIQPKSVRLVYANTRD----- 530
Db 2277 DDINDNPPVFAQQSYAVTLSEASVIGTSVQVVRATDSSEPNRGISYQMFNGHSHKSHDHF 2336
Qy 531 -----GPAKGKDVVQSGDITLPVPGADPRKLVAFSDVLSGQAHVLSVATKVTWCW 581
Db 2337 HVDSSSTGLISLLRTLDEYQSRQHTIFVRAVDGGMPTLSSDV-----IVTVDTDLNDN 2389
Qy 582 PNLGRGFGQPIITLPGFSQAPATE---FNPAQVYLAD-----LDG----- 617
Db 2390 PPL-----FEQIYEAREISEHAPGHGFTVCYKAYDADSSDDKLOYSTLSGNDHKHFVIDS 2445
Qy 618 -SGPTDLIYVHTNRDLDFLNKSGNGPAPVTLRPEGLRFDHTCQLOMADVQGLGVASLI 676
```

Db 2446 ATGIIITLNLHRHAKPYS-----LNLVSVDGV-FRSSTQVHVTVIGG-----N 2489

Qy 677 LSVPHMSPHHRCDI/TNNKP--WLLNEMNNMG-----VHTLYRSSSQFWLDEKA 726

Db 2490 LHSAPFLQNEVEVELAENAPLHLLVMEVKTITDGDGSIYGHVTHYI/VNDFAKDRFVINERG 2549

Qy 727 AALT-----GQTPVCVLPFPPIHTLWQTETEDISGNKLVTT 763

Db 2550 QIFTLKLDRETPAEKVSVRLMAKDAAGKVAFC-----TVNVILTDDNDNAFQFRAT 2602

Qy 764 LRY-----ARG-----AWDGRERERFGVVEQTDHQLAAGNAPERTPPALTKN 808

Db 2603 -KYEWIGSSAAKGTISVVKVLASDADEGNSADITVAIEADSESVKENLEINKLSGVITTK 2661

Qy 809 WYATGLP-----VIDNA-----LSTEYWRDDQAFAGSPRFTTWQDNKQVPL 850

Db 2662 ESLIGLENEFFTFVRAVDNGSPSKESVVLVVKILPPEMQLPKFSEPFYFTVSEDVPI 2721

Qy 851 TPEDDNRWYFNRAKGLQLLSELYGLDDSTNKHVPYVTFPRSQV-----RRLQHTDSRYP 907

Db 2722 GTEIDLIR-----AHSQGTVLVSLVKGNTPESNRDESFVIDRQSGRLKLEKSLDHETTKW- 2776

Qy 908 VLMSVVSZNRVHYERIASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQQPAINLYP--- 964

Db 2777 -----YQFSILARCTODDHEMVASVD-----VSIOVDANDNSPVFESSPYEA 2819

Qy 965 ---DTLPDKLLANSYDDOQROLRLTYQQSSWHHLTNTNVRVLGLPDPSTRSDIFTYGAENV 1021

Db 2820 FIVENLFG-----GSRVIQIRASDADSG-----TNGQV-MYSLDOSQSVESFALNM 2867

Qy 1022 PAGGLNL--ELLSQKN-----SLIADKPREVLGQOKTAYTDGQNTPLQTFTRQALLAF 1074

Db 2868 ETGMITTLKELDHKRDNYQIKWASDH-GEKIQLSSTAI/VDVTVDVNDSPPR----- 2920

Qy 1075 TETTVFNQSTLSAFNGSIPSDK-----LSTT---LEQAGYQOQTNVLPRTGEDKWV 1123

Db 2921 -----FTABIYKGTVSDEDDPOGGVITAILSTTDADSEENRQVTFPI---TGGDPL-- 2967

Qy 1124 AHHGYTDYGTAAQFWR-----POKOSNTQLTGKITLIWDANYCVVVQTRDAAGLTS 1175

Db 2968 ---GQFAVETTQNEKVVVKKPLDREKDNVILT--IT-----ATD 3003

Qy 1176 AKYDWRFLTPVQLTDINDNOHL-----ITLDALGRPITLRFWGT-----NGKMTG 1221

Db 3004 GTFSSKAIVEVKVLANDNSPCKEYKLYSDTIPEDVLPCKLIMOISATDADIRSNABEIT- 3062

Qy 1222 Y-----SSEPKASFSPSDVNAIEELKPLPVACQVYAPESWMPVLSOKTFNRLAQDWQ 1277

Db 3063 YTLGSGAEKFKLNDPTG-----ELKTSTPL-----DREEQ 3093

Qy 1278 KLYNARIITED--GRICTLAYRRWQSOAKIPLQILSLN-NGPRL--PPHSL----- 1324

Db 3094 AVYHLLVRATDGGGRFC-----QASIVLTLEDVNDNAPEADSPAYITVENTEP 3143

Qy 1325 -TLTTRDYDHPPEQIROQVWFS-----DG-----FGRLLQAA-----A 1357

Db 3144 GTLLTRVQATADAGLNKILYSLDSADGQFSINELSGIILEKPLDRELQAVVTLSLK 3203

Qy 1358 RHEAGMARQNEGSLINVOHTENRWAVTGRTEYDNKGQPIRTQPY 1405

Db 3204 AVDQGLPRRLTATGTVIIVSVLDI-----NDNPPVFYREY 3238

RESULT 7

US-11-113-424-53
 ; Sequence 53, Application US/11113424
 ; Publication No. US20050260713A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gangolli et al.
 ; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
 ; FILE REFERENCE: 21402-225
 ; CURRENT APPLICATION NUMBER: US/11/113,424
 ; CURRENT FILING DATE: 2005-04-21

; PRIOR APPLICATION NUMBER: 60/256,704
 ; PRIOR FILING DATE: 2000-12-19
 ; PRIOR APPLICATION NUMBER: 60/311,590
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: 60/257,314
 ; PRIOR FILING DATE: 2000-12-20
 ; PRIOR APPLICATION NUMBER: 60/311,613
 ; PRIOR FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: 60/315,617
 ; PRIOR FILING DATE: 2001-08-29
 ; PRIOR APPLICATION NUMBER: 60/307,506
 ; PRIOR FILING DATE: 2001-07-24
 ; PRIOR APPLICATION NUMBER: 60/322,358
 ; PRIOR FILING DATE: 2001-09-14
 ; PRIOR APPLICATION NUMBER: 60/294,075
 ; PRIOR FILING DATE: 2001-05-29
 ; PRIOR APPLICATION NUMBER: 60/288,153
 ; PRIOR FILING DATE: 2001-05-02
 ; NUMBER OF SEQ ID NOS: 190
 ; SOFTWARE: Patent in Ver. 2.1
 ; SEQ ID NO 53
 ; LENGTH: 2515
 ; TYPE: PRT
 ; ORGANISM: Drosophila melanogaster
 ; US-11-113-424-53

Query Match 1.8%; Score 139; DB 7; Length 2515;
 Best Local Similarity 16.5%; Pred. No. 0.11;
 Matches 236; Conservative 161; Mismatches 446; Indels 586; Gaps 59;

Qy 1 MONSDPSITELSLPKGGGALTGMGEALTTPGP-----DCMAALSPLPISAGRGVAPAPT 56

Db 1027 VRDNDYSQPELNWE-----AVVGSGERCLPDEAHCGDALAKLAPYKGIATSSONI 1082

Qy 57 LNVNSGA-----GNSPFGLW-----DCNVMTIRRRTHFGVPHYDE 92

Db 1083 LYFADGNTIRMVDRDGVSTVLGNHMKSHKWPICPGSTLKEMLHWPTELAVSPMDN 1142

Qy 93 TDTFLG-----PEGEVLVADQPRDESTLOGINLGATFTVTGYRSLRSHFSRLEYW 144

Db 1143 TLHIIDHMLRMTDPGRVVRVISGRP-----LHCATASTAYDTDLATHATLV--- 1189

Qy 145 QKTTGKTDFWLIYSPDQGVHLLGKSPQARISNPSQTTQTAQWLLLEASVSRGEIYYQY 204

Db 1190 MPQS-----IARFPLGELY-VAESDSQRINRVVIGTGRGIAPFAGAESKCNCL--- 1237

Qy 205 RAEDDTGCEADEITHLQATAQ-----RYLHVYVGNRTASETLPLGLDSAPS 252

Db 1238 ----ERGDCDCEAEHYLATSAKFNTIAALAVTPDSSHVHIADQANVIRSVSVSSIPASPS 1293

Qy 253 ----QADMLF-----YLVFDYGBRSNNLKTTPPAFTSTTGSWLC 285

Db 1294 REVEIYAPDQGEIYIFNRFQGHVSTRNLLTGETTVVFTYVNTVNSGKLSVTVDAGNKVF 1353

Qy 286 RODRFSRYEYGFETRRLCLR---QVLMYHHLQALDS-KIT-EHNGTILVSR----- 332

Db 1354 LLRDYTSQVNSIENTKGOKCLRMTRMKMLHELSTPDNPNVNTYEHGPTGLLRTKLDSTG 1413

Qy 333 --LILNLYDE---SASTAST-----LVFVVR- 350

Db 1414 RSYVNYDEFGRLTSAVTPTGRTVIELSPDLSVKGQVKSNAQKEMSLLTQCATVVRN 1473

Qy 351 ----RVGHEQDGNVVTLP-----LELAYQDFSPRHHAHWQPMQMDVLANFNAIQRWQLV 399

Db 1474 GAESRITVDMGSGTTSITPWGHNLOMEVA-----PYTILA-----EGS 1512

Qy 400 DLKGEGLPGLLYQ-----DKGAWWY--RSAQRL-----GEIGSDA--- 432

Db 1513 PLLGESVVPKQRTETAGDLANRFEMRYFVRROQPLQAGKSKGPPRPVTEVGKLRVN 1572

Qy 433 -----VTWEK-MQPLSVIP----- 445

Db 1573 GDNVLTLEYDRETSQSVVMVDDKQELLNVYDTRSPISFRPQSGDYAYVDLEYDRFGLR 1632

```
Qy 446 -----SLQSNASLVGIN-GD-----GOLDWVITGPGGLRGVHSORPD 480
Db 1633 VSWKGVLEAYSPDRNGRLNEIKYGDSGTWVYAFKDMFGSLPKVITPRRSYLLQVDD 1692
Qy 481 G-----SWTRFTPLNALPVE-----YTHPRAQ 502
Db 1693 AGALQSLTTPRGHTHAFSLQTSGLGFKYQYVSPINRHPFELYNDEQILLAKIHPHOSGK 1752
Qy 503 LADLMG-AGLSDLVILGPKSVRLVANTRDGPAGKDVVQSGDITLVPFGADPRKLVAF-- 559
Db 1753 VAFVHTAGRLTETLAGLSSTHYTYQDTTSLVKSVEQVE-----PGFELREFEKYHA 1804
Qy 560 -----SDVLGS-----GOAHL-----VEVSATKVTCPNPLGRGRFG 590
Db 1805 GILDEKLRFGSKSLASARYKYAYDGNARLUSGEMAIKDLKELPTTRYKYSQNGQLLEV 1864
Qy 591 Q--PITLPGFSQAPATEFNPQVY--LADLDGSGPTDLIVVHTNRIDF---LNKSGNGFAE 644
Db 1865 QDLKITRANFNRTVIQDSAKQFFAIVDYDQGRVKSVMNVKNIDVFRLELDYDLNRK 1924
Qy 645 PVTLRPFEGRLFD---HTCQLQMAADVQGLGVASLILSVPHMSPHHWRCDLTNNKFWLLNE 701
Db 1925 SQKTTFCRSTAFDKINTYNADGHVVEVLG-----TNNWKYLFDE 1962
Qy 702 MNNMGVHHTLRYRSSSQFWLDEKAAALTTGQTPVCVLPPEPIHTLMQTEDEISGNKLV 761
Db 1963 NGNTVG-----VDQGEKFNLYDI-----GDKVI 1987
Qy 762 TLLYARGAMDR-----EREPGRGYVEQTDSHQLAQNAPERTPPALTKWYATGL 814
Db 1988 KVGDFENNYDARGFVVKRGQYR-----YNNRGQLIHSFERER-----FQSM----- 2031
Qy 815 PVIDNALSTEWDRDQAPAGESPFTTQDNKDVLPDTPEDDNSRYWENRALKGQLLSEL 874
Db 2032 -----YYDDRS-----RLVAHDNK-----GNTQYY----- 2055
Qy 875 YGLDDSTNKHVPYVT-----EFBSQVRRLOHTDSRYPLWSSVVESENY 919
Db 2056 -----ANRTPHLVTHVFPKISRTMKLFYDDRDMIALHEHQRYV----- 2098
Qy 920 HYERIASDQCSQNTLSSDRFGQPL-----KQLSVQYPRRQOPAINLYPDTLPDKLLANS 975
Db 2099 -----ATDQNGSPLAFFDQNGSIVKEMKRTPGRIIKOTKPEFFVPID 2141
Qy 976 YD-----DQORQLRITYQSSWHHL-----TNNTVRVGLPDSTRSDIFTYGAEVNPAGGLNL 1028
Db 2142 FHGGLIDPHTKLVTYEQRYDPHYGQWMTPLWETLATEMSHTDVFYRYHN----- 2193
Qy 1029 ELLSDKSLIADDPREVL-----GQOKTAYTDQONTTP 1062
Db 2194 -----NDPINRPNQNTMIDLSWLQIFGYDLNMQSSRYTKLAQYTP 2236
```

RESULT 8

```
US-11-113-424-14
; Sequence 14, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
```

PRIOR FILING DATE: 2001-08-29
PRIOR APPLICATION NUMBER: 60/307,506
PRIOR FILING DATE: 2001-07-24
PRIOR APPLICATION NUMBER: 60/322,358
PRIOR FILING DATE: 2001-09-14
PRIOR APPLICATION NUMBER: 60/294,075
PRIOR FILING DATE: 2001-05-29
PRIOR APPLICATION NUMBER: 60/288,153
PRIOR FILING DATE: 2001-05-02
NUMBER OF SEQ ID NOS: 190
SOFTWARE: Patent in Ver. 2.1
SEQ ID NO 14
LENGTH: 2769
TYPE: PRT
ORGANISM: Homo sapiens
US-11-113-424-14

Query Match 1.7%; Score 136; DB 7; Length 2769;
Best Local Similarity 18.2%; Pred. No. 0.21;
Matches 332; Conservative 211; Mismatches 563; Indels 714; Gaps 101;

Qy 66 SPFGLGHCNVMTIRRTTHFGVPH-----YDETDTFLGPEGEVLVVADQP----- 110
Db 866 NPLCLG-SPNPLDIIQETQVPSQONLHSHFYDRIKFLVGRDSTHIIIPGENPFDGHCACVI 924
Qy 111 -----RDESTLQGINLG-ATFTVTGYR-SRLESFHS-----RLEYMQPKTKG 151
Db 925 RGQVMTSDGFLVGVNLSFVNPNLFGYTI SRDGSFDLVTNGGISIILRPER-APFITQE 983
Qy 152 TDFWL-----IYSPDQVHLLGKSPQARISN-----PSQTTQAO----- 186
Db 984 HTLWLPWDRFVWETIIMRHEENEIPSCDLSNFARNPNVPSPLTSFASSCAEKGPVP 1043
Qy 187 --WLEASVSRSRGEIYYQRAEDDTCEA---DEITH-----HLOATAQRYLHI 231
Db 1044 EIQAQBEIISGCKMFLSYLSERTPGYKSVLRISLTHPTTIPFNLKMLVHLMVAVEGRLEFR 1103
Qy 232 VYGNRTASETLPGLDGSAAPSQADWLFYVFD---YGERSNLKTTPAFSTTGSWLCRQ 287
Db 1104 KWF-----AAAP---DLSYFFWTKDVTYVQKVFGLS---EAFVSVG----- 1139
Qy 288 DRFSRYEY-----GFEIRTRLCROVLMYHHLQALDLSKITEHNG--- 326
Db 1140 -----YEYESCPDLILWEKRTTVLQGYEIDASKLGGWSLQKHALNIQSGIL-HKNGEN 1193
Qy 327 -----PTLVSRILN-----YDESALASTLVFVR 351
Db 1194 QFVSQPPFVIGSINGNGRRRSISCPSCNGLADGNKLLAPVALTCGSDGLYVGDFTVIRR 1253
Qy 352 VGHEQDGNVVTLPPELEAYQDFGPRH---HAHQPMQMDVLANFNAIQRWLVDLKGGLPG 408
Db 1254 I--FPSGNVTNI--LELRKDFRHSHPAKYIYATDPM----- 1289
Qy 409 LLYQDKGAWYRQAQRLGEIGSDAVTWKMQPILSVIPSLQSNASLVLDINGQGLDWITG 468
Db 1290 -----GAVFLSDSNSRRVF--KIKSTVWVKDLVKNSEVWAGTQDCLPFDTR 1335
Qy 469 PGLRGYHSQRPDGSWTFPTLNALPVEYTHPRAQLADLMGAGLSDLVLIGPKSVRLVANT 528
Db 1336 CG-----DGG-----KATEATLTNPRGTIVDKFGL-----IY--- 1362
Qy 529 RDGFAGK---KDVVQSGDITLVPFGADPRKLVAFSDVLGSGQAHLVEVSATKVTCPNPLG 585
Db 1363 ---FVDGTWIRRIDQNGII-----STILGSDNL---TSAPLSC----- 1395
Qy 586 RGRFGQPIITLPGFSQAPATEFNPQVYLAIDLGGSGFTDLIYYVHTNRIDIFLNKSGNGFAEP 645
Db 1396 -----DSVMDISQVRELEW-PTDLAINPMDNS---LYVLDNN----- 1427
Qy 646 VTLRPFEGRLFDHTCQLQMAADVQGLGVASLILSVPHMSPHHWRCDLTNNKFWLLNEMNN 705
Db 1428 VWLQISE---NH---QVRIVAG-----RPMH---CQVPGIDHFLUSK----- 1460


```
Db 663 QLGH-LTAHGCTNVVAFSFMNGMALPN-D-YLISKDVTGATKVKYXYAVNDGFFGDHYAV 720
Qy 634 FLNKSNGFAFPVTLRFPBGLRFPHTCOLQADVQGLVASLIILSVPHMSPHWRCDLTN 693
Db 721 MISKTGTN-AGDFTVVFEE-----TPN-----741
Qy 694 MKPMLNEMNNMGVHHTLRYSRSSQFWLDEKAAALTTGTFVFCYLPFPPIHILWQETED 753
Db 742 -----GIN-----KGARFGLSTEA-----NGAKP-----QSVWIERTV 771
Qy 754 EISGNKLVTLRYARGAWDGRBRFRGYYVEQDSSHQLAQGNAPERTPPALTKRWYATG 813
Db 772 LPAGTKVAFRHY-----NCSDLNYI-LDDIQFTMGSGP--TPDITYTVYRDG 818
Qy 814 LPVIDNALSTYEWDDQ-----PAGESPRTTWQDNKDVPPLPEDDNSRYWF 861
Db 819 TKIKEGTETTFEEDGVATGNHCEVKYTAGVSPK-----ECNVTVDPVQNPQNL 873
Qy 862 NRALKGQLLRSELYGLDDSTNKHVPYTVTEFRSQVRLQHTDSRYPLWSSV-----913
Db 874 TGSVAGQKVTLK-WDAPNGTNPNGTTLSES-----PENGIPASWKTIDADGDN 925
Qy 914 -----VESRNY-HYERIASDPQ-----929
Db 926 WTTTPPGGTSFAGHSAICVSSASYNFE---GPQNPONLYLTPBELSLPNGGTLTFW 981
Qy 930 CSQNTLSSDRFGQPLKQLSVQYPRQOPAINLYPDILPKLLANSYDDQORQLRTYQ 989
Db 982 CAODANYASEHY-----AYASSTGNDASNFANALLEEVLTAKTVTVAPEAIRGRVQ 1034
Qy 990 SSWHLTNNTVRVLGLPDSTR-----SDIFTYGAENV--PAGGLNLELLSDKNSLI 1038
Db 1035 GTWTQKT-----VOLPAGTKVAFRPHGCTDFWNLNDDVEIKANGRADFTTFESST 1088
Qy 1039 ADDRPREYLGQOKTAYTDQNTTLPOTPTROALIAFTETTV-----FNQSTLS 1086
Db 1089 HGEAPAEW--TTIDADGQGWCLSSQQLDMLTAHGCTNVVASFNGMALNPDNYLIS 1146
Qy 1087 -----AFNGSIPSDKLSLTLEQAGYQOTNY--LFPPT--GEOK-----1120
Db 1147 KDVGTATKVKYXYAVNDGFFGDHYAVMISKGTNAGDFTVVFETPNNGKGGARFGLST 1206
Qy 1121 -----VW-----VAHGY-----TDY-----1131
Db 1207 EANGAKQSVWIERTVDLPACTKVAFRHVCNSDLNVLILLDDIQFTMGSGPTDYTYTV 1266
Qy 1132 ---GTAAQFWRPQKQSNLTQLGKITLIWDANYCVVQTRDAAGLTTSKADWRFLLTPVQL 1188
Db 1267 YRDGTKI-----KEGLTETTFEEDGVATGNHCEVKYTAGVSPKECVNVT--VDPVQF 1319
Qy 1189 TDINDNQHLITLDALGRPITILRFPWGTENG 1217
Db 1320 NPVQN-----LTGSVAGQKVTLK-WDAPNG 1343

RESULT 10
US-11-052-554A-174
; Sequence 174, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052.554A
; CURRENT FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 174
```

```
; LENGTH: 2204
; TYPE: PRT
; ORGANISM: Porphyromonas gingivalis W83
US-11-052-554A-174

Query Match 1.6%; Score 127.5; DB 7; Length 2204;
Best Local Similarity 18.8%; Pred. No. 0.67;
Matches 273; Conservative 175; Mismatches 488; Indels 517; Gaps 74;

Qy 105 VVADQPRDESTLOGINLGATFTVTGYRSRLSHESFSLRYEQPKTKGTDFWLIYSPQGV 164
Db 1027 IVADQSQAGSNGIYSV-----SALNVNRPILY-----TNKPSIALV-NQSGEV 1068
Qy 165 HLICKSPQ--ARISNPNQTTQTQAWLLEASVSRGEQIYYQYRAEDDTGCEADBITHLQ 222
Db 1069 ELVGKTEGWLKRLSNPSSATAPYVWLALPTS-----GLTIEKVT--D 1109
Qy 223 ATAQRYLHIVYVGNRTASETLPLGLD-GSASQADWLFLV-FDYGERSNLKTTPAPSTT 280
Db 1110 AAGTEMAFTTYSGGKMYRLSEAGVPVGS-----LDYTHFTYSGCS-----PIALKAM 1158
Qy 281 GSWLCRODRSRVEYGFETRLCRQVLMVHHLQALDS--KITEHNGP-----TLVSR 332
Db 1159 GWMNCSAYPLSDEY-----VCSSQVLDLKLPLPAAMELTEIAVDPDTAAATLCST 1210
Qy 333 L-----ILNYDESAIASTLVFRRVVGHEQDGNVVTLPPELAYQDFSPRHHAHQMDVL 387
Db 1211 LEYIYSIQSTDNANVYSPFESI--FPSEGLVTPNQVQVEY-----PAGSGNWAALNV 1262
Qy 388 AN-FNAIQRWQLVDLKGEGPLGLLYQDKGAWWYRSQRLGEIGSDAVTWEMQPLSVIPS 446
Db 1263 NNSVNLLOHPALTTI-----GYLKLKKEGSEN-----DNQRKILVKFY 1300
Qy 447 LQSNASLVD-----INGDQGLDWMVITGPGLRGVHSQRPDGSWTFP--TP--LNALPVEYT 497
Db 1301 IKTECFVSQGNFRVRADG-----RNACNQNAKSGSLAISTPIRINGAIBPYT 1349
Qy 498 HPAQLADLADLADLADLADLADLADLADLADLADLADLADLADLADLADLADLADLADLAD 557
Db 1350 --TSASTQLVTTTTSQSDCKAPKVKVQVTVGGETTPKAYL-----EITLP-----L 1395
Qy 558 AFSDVLGS---GQAHLEVEVSATKVTCPNIGRGFRGQPIITLPGFSQPAITFNPQVYLAD 614
Db 1396 GFKYVTGSYAPDNTHPGVGNASPAQT-----EEVLTITANGEDIKIN-----VKAG 1441
Qy 615 LDGSGPTDLLVHTNRDLIDFLNKSNGFAFPVTLRFPBGLRFDHTCO-LQWAD---VOGL 670
Db 1442 LT-SGQSFAYTLEMKEDDDNVPCAGNHTIELVNVVEIEGL-----WCSGVQCAETLVV 1496
Qy 671 GVASLIILSVPHMSPHWRCDLTNNKPLNEMNNMG--VHHTLRYRSSQFWLDEKAAA 728
Db 1497 NKPEFELDKPYL-----DIT-----VISAVSTFSGKENLTIEYKVN-----1534
Qy 729 LTTGQTPVCYLPPIHTLWQETEDEI--SGNKLVTTLRYARGAWDGRERFRGFGVVEQ 786
Db 1535 -TSTTQPL--KPGAVVTLFSDKONNVFSGDVAVAQELV-----AEI 1575
Qy 787 TDSHQLAQGNAPERTPPALTKWYATGCP-----VIDNALSTYWRDDQAFAG 834
Db 1576 TNPPLTQIMKVGKGVSSHTGNLVLITLPKDCGYCEIKSPMVTNLNHLPSNYW-----IGG 1630
Qy 835 FSPRFTTWQD-----NKDVPILTPEDDNSRYWFNRALKGQLLRSELYGLDDSTNKHVPYTV 889
Db 1631 TVGKPNWKEKFPNNWINDQVDAED-----1655
Qy 890 TEFRSQVRRLQHTDSRYPLWSSVVESRNHYERIASDPQCSQNTLSSDRFGQPLKQLS 949
Db 1656 VEFATEVN--NPTDPNPN---KSGPAKENLHLDI-----HQNGTAG-----1692
Qy 950 VQYPRROQPAINLYPDILPKLLANSYDDQORQLRTYVQSSWHLTNNTVRVLGLPDST 1009
Db 1693 -----RVINGNLNDSKDLVIT-----TGNTQTINGVVEEDN 1723
```


Db 1115 ---TLVQNPLLLLENKPSVEKTP-TIKYFRTEISQVTPPTGAVMTYAPTSIPMEKTHKVN 1170
Qy 919 YHYERIASDPQCSQNTILSSDRFG---QPLKQLSV-QYPRQOPAINLYPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSSLSGAIKTPPMIIITAITFRSRKIPW-----QQNFV 1221
Qy 974 NSYDDQOR---QLRLTYQQSS-----MHHLTNNTVRLGLP-----1006
Db 1222 NNHPKGLRNLQHKVSLQKSTAVMLPKTSPALPQKQSLPSHHTTKTHNPGSLPKKELP 1281
Qy 1007 -----DSTRSDIFTYGA-----ENVPAGLNLLELLSDKNLSLIADDKPREY 1046
Db 1282 PPLNPMPLPSIISKDSSTKSIISTQAIPTATPTTPPASVITYETQTERSRAQTQREOE- 1340
Qy 1047 LGOQKATWDQONTPLQ-----TPTRQA--LIAFTETTVNQSTLSAFNGSIPS DK--L 1097
Db 1341 --POKKNRTD-PTNSPDQSGFTTPTAMTPVLTATTSV--KPSVSATFHTSPENTTGI 1395
Qy 1098 STTL 1101
Db 1396 SSTI 1399
RESULT 12
US-10-453-372-716
; Sequence 716, Application US/10453372
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsbrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: Curaseqlist version 0.1
; SEQ ID NO 716
; LENGTH: 2602
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-716
Query Match 1.6%; Score 127; DB 6; Length 2602;
Best Local Similarity 19.1%; Pred. No. 0.95;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;
Qy 91 DETDTFLQPEGEV-----LVWADQPRDEGTL-----QGINLGATFTVTGYRSLRLESH 137
Db 309 DSSSAFTSPQGFAPFGSLTNWTDQSGNEANVCSIQKPSRTSPFIATFENDIVLNTS 368
Qy 138 FS-----RLEY-----WQPKTKTKDFWLIYSPDQV-----HLLGKSPQARIENPSOTT 182
Db 369 FSTFLVCNIDYGHIQPVNQ-----ILALYSDSPILERSHLLSETPQL-----411

Qy 183 QTAOWLLEASVSRGEQIYYQYRAE---DDTGCEADBITHLOQATAQRYLHIVYGNRTA 239
Db 412 ---YYKYKQVAPKPEIDFTNIEADLRADPSWLMQDQISLOLNRTATFTFLQIQYSSDA 467
Qy 240 SETLPGLDGSAQSADMLFVLFVDFYDGRSNNLK---TPPAPSTTG-----SW 283
Db 468 QITLPRAB-MRPVHKWTMI-----SRDNNTKLEHTVLVGGTVGLNCPGQGDPTPHVDM 520
Qy 284 LCRDRFSRVEYGEF-----IRTRLCRQVL-----MYHHLQALDSKITEHNGPTLVS 331
Db 521 LLADGSKVRAPYVSEDGRILIDKSGKLEQWADSFDTGVVHCIS-----565
Qy 332 RLILNYSATASTLTVFRRVGHEDGNVVTLPLELAYODFSRHHAAH--QPMVDVLN 389
Db 566 ---NYDDADILTY-----RITVVELEVEAYQE-NGIHHTVFIGETILDLPCH 607
Qy 390 FNAIQ-----RWQLVDLKGEGLP--LLYQDKGAWMYRSAQRLGIGSDAVTWKMQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRILOV 646
Qy 444 IPSLOSNASLVDINGDQLDWITGPGLRGVHSORP---DGSWTRFTPLNALPVEYTHPR 500
Db 647 TPKDQGYRYRCVAANPSG-VDFLIFQVSVK-MKGQRPLEHDE-----TEGSGLDENPI 698
Qy 501 AOLADLMGAGLSDLVLI-----GPKSVRLYANTRDGF---532
Db 699 AHLKEPPGAQLRTSALMBAEVGKITSSTKKNYRELTLQRRGDSHRRFRHPPS 758
Qy 533 -----AKGKVVQSGDITLPPVGADPRKLVAFSDVLG-----SQOAHLVE 572
Db 759 ARRIDPQHWAALEKAKKNAMPDKRENTVSP---PPVVTQLPNIPEGEDSSGMLALHE 815
Qy 573 ---VSATKVTCPNML-----GRGFGQIPITLPGSQPATEENPAQVILADLDGSGPTDL- 623
Db 816 EFMVPATKALNLPARTVTDASRTISDSPTMNINY---GTEFSPV-VNSQILPPEPTDFK 871
Qy 624 --IYVHTNRLDIFLNKSGNGFAEPVTLR-----PP-----EGLRFDH-----658
Db 872 LSTAIKTAMSKNINPTWSSQIQGTTNQHSTVFPLLLGATEFQDSQDMGREHFQSRP 931
Qy 659 --TCOLQWADYQGLGVAS-----LILSVPHMSPHHWRCDLTNMKPLLNNMNNNGVHHT 711
Db 932 PITVTRTIKDVNVKMLSSTTNKLLLESVNTNSHQ-----TSVRE--VSEPRHNHFYSHT 984
Qy 712 LRYSSSQFQWDEKAAALITGQTPCVLPPIIHLWQETETEDSISGNKLYTTLAYARGAW 771
Db 985 TQILSTSTFPSPDPTAA-----HSQFP-----IPRNSTVNIPLFRR---1020
Qy 772 DGREREPFGYVEQTDSDHQLAQGNAPERTPPALTKNMYATGLPVIDNALSTEYWRDQA 831
Db 1021 FGORKIGGRRI-----ISPYRT-PVLRHRHYSI-----FRSTTRGSSEKS 1061
Qy 832 FAGFS-----PRFTTWQ-----DNKDVPLTPEDDNSRYWFNRA 864
Db 1062 TTAFSATVNLVNTCLSPRERLTATAALSFPSPAAPITFPKADIARVPSEST-----1114
Qy 865 LKGQLLRSELGLDD--STNKHVITYTFRSQVRRLQHTDSRYPLWSSVVSER---N 918
Db 1115 ---TLVQNPLLLLENKPSVEKTP-TIKYFRTEISQVTPPTGAVMTYAPTSIPMEKTHKVN 1170
Qy 919 YHYERIASDPQCSQNTILSSDRFG---QPLKQLSV-QYPRQOPAINLYPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSSLSGAIKTPPMIIITAITFRSRKIPW-----QQNFV 1221
Qy 974 NSYDDQOR---QLRLTYQQSS-----MHHLTNNTVRLGLP-----1006
Db 1222 NNHPKGLRNLQHKVSLQKSTAVMLPKTSPALPQKQSLPSHHTTKTHNPGSLPKKELP 1281
Qy 1007 -----DSTRSDIFTYGA-----ENVPAGLNLLELLSDKNLSLIADDKPREY 1046
Db 1282 PPLNPMPLPSIISKDSSTKSIISTQAIPTATPTTPPASVITYETQTERSRAQTQREOE- 1340

QY 1047 LQOQKATYDQNTTLPQ-----TPTRQA--LIAFTETTVFNQSTLSAFNGSIPSDK--L 1097
Db 1341 --PQKNRTD--PNISPDQSSGTTTPTAMTPPVLTAAETSV--KPSVSATHTSPPTTGTI 1395
QY 1098 STTL 1101
Db 1396 SSTI 1399
RESULT 13
US-10-453-372-666
; Sequence 666, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 666
; LENGTH: 2617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-666
Query Match 1.6%; Score 127; DB 6; Length 2617;
Best Local Similarity 19.1%; Pred. No. 0.96;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;
QY 91 DETDFLGEV-----LVVAQPRDESTL-----QGINLGAFTVTCYRRLSH 137
Db 309 DSSAFISPGFMPPGSLTNMTQSGNEANMVCSIQKPSRTSPIAETENDYVLNTS 368
QY 138 ES-----RLEY-----WPKTKTGDWLIYSPQGV-----HLGKSPQARISNPSQT 182
Db 369 FSTFLVCNIDYGHIPVWQ-----ILALYSDSPILERSHLLSETPOL----- 411
QY 183 QTAQWLLASVSRRGEQIYYQVRAE---DDTCEADEITHLQATAQRYLHIVVYGNRTA 239
Db 412 ----YKQYQVAKPEDIFTNEADLRADPSMLQDQISLQLRNTATTFSTLIQYSSDA 467
QY 240 SETPLGLGSAQSADWLFYLVFDYGRSNNUK---TPPAFSTTG-----SW 283
Db 468 QITLPRAE--MRPVGHKWTMI-----SRDNTKLEHTVLVGTVGLNCPCGQGDPTPHVDW 520
QY 284 LCRQDRFSRYEYCFE-----IRTRLCQVQL-----MYHQLQALDSKITEHNGPTLVS 331
Db 521 LIADGSKVRAPVSDGRILIDKSKGLQMAADSPDTGVYHICSS----- 565
QY 332 RLILNYDESATILVFRVRGHEQDGNVVTLPPLPELAYQDFSPRRHAW--QPMVDVLAN 389

Db 566 ----NYDDADILTY-----RITVVEPLVEAYQB--NGIHHTVTFIGETLDDLPCB 607
QY 390 FNAIQ-----RQVLVDLKGEGLPG--LAVQDKGAWMYRSAQRLGEIGSDAVTWKMQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRILOV 646
QY 444 IPSLQSNASLVYDINGDQLDWITGPGRLRGVHSQRP---DGSWTRFTPLNALPVYTHPR 500
Db 647 TPKDQGYRVCVAANPSG--VDFLIQVSVK--MKGQRPLEHDGE-----TEGSLDESNP 698
QY 501 AOLADLMAGLSDLVLI-----GPKSVPLIYANTYRDFG--- 532
Db 699 AHLKEPPGAQLRTSALMEAEVKGHTSSTSKHNYRELTLQRRGDSHRRFRNRHFPSP 758
QY 533 -----AKGKDVVQSGDITLPPVGADPRKLVAFSDVLG-----SGQAHLVE 572
Db 759 ARRIDPQHWAAALLEKAKKNAMPDKRENTTVP---PPVVTQLPNIPGEDSSGMLALHE 815
QY 573 ---VSATKVTCWPNL-----GRGRFGQITILPGFSQPATEFNPAQVYLADLDGSGPTDL- 623
Db 816 EFMVPATKALNLPARTVTADSKTISDSPTNINY---CTEFPSPV--VNSQILPPEEPTDFK 871
QY 624 --IVHTWRLDI FLKSGNGFAEPVTLR-----PP-----EGLRFDH----- 658
Db 872 LSTAIKTTAMSKNINPTMSSQIQGTTNQHSSTVPFLLLGATEFQDSQDQGRHFPQSRP 931
QY 659 --TCOLOMADVOGLGVAS-----LILSVPHMSPHHWRCDLTNMKPWLNNMNNMGVHHT 711
Db 932 PITVTRMIKDVNVKLSSTTKLLESVNTNSHQ-----TSVRE--VSEPRHNFYSHT 984
QY 712 LRYRSSQFWLDEKAAALTTGQTPVCYLPFPIHTLMOTETEISGNKLVTLRLVARGAM 771
Db 985 TQILSTSTFPSPDPTAA-----HSQFP-----IPRNSTVNIPLFR--- 1020
QY 772 DGRERFEGYVEQTDHQLAOGNAPERITPALTKNWYATGLPVIDNALSTEYWRDQA 831
Db 1021 FGRQKIGGRGRI-----ISPYRT--FVLRHRYSI-----FRSTTRGSSSKS 1061
QY 832 FAGFS-----PRFTTWQ-----DNKDVPLTPEDDONSRYWFNRA 864
Db 1062 TTAFSATVNLVTCLSCLPRELITATAALSFPSPAIPFPKADIARVSEEST----- 1114
QY 865 LKGQLRSELYGLDD--STNKHVPYVTFEFSQVRRLQHTDSRYVPLWSSVVSER---N 918
Db 1115 ---TLVQNPLLLLENKPSVEKTP--TIKYFRTEISQVTPGAVMTYAPTIPMEKTHKNV 1170
QY 919 YHYERIASDPOCSQNTLSSDRFG-----QPLKQLSV--CYPRQOQAINLYPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSLSGAITKPPMTIITAITFRSRKIPW-----QQNFV 1221
QY 974 NSYDDOQR---QLRLTYQSS-----WHLTNTNVTVRVLGLP----- 1006
Db 1222 NNHNPGRKLRNQHKVSLQNSLOKSTAVMLPKTSPALPQOQSLPSSHHTTKTNPGSLPTKXELP 1281
QY 1007 -----DSTRSDIFTYGA-----ENVPAGLNLLELSDKNLSIADDPREY 1046
Db 1282 FPPLNPLMPSIISKDSSTKSIISITQTAIPATTPFPASVITVETOTERSAQTIQREQE- 1340
QY 1047 LQOQKATYDQNTTLPQ-----TPTRQA--LIAFTETTVFNQSTLSAFNGSIPSDK--L 1097
Db 1341 --PQKNRTD--PNISPDQSSGTTTPTAMTPPVLTAAETSV--KPSVSATHTSPPTTGTI 1395
QY 1098 STTL 1101
Db 1396 SSTI 1399

RESULT 14
US-10-453-372-732
; Sequence 732, Application US/10453372
; Publication No. US20060003323A1
; GENERAL INFORMATION:

```

; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraseqList version 0.1
; SEQ ID NO 732
; LENGTH: 2617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-732

Query Match          1.6%; Score 127; DB 6; Length 2617;
Best Local Similarity 19.1%; Pred. No. 0.96;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;

Qy 91 DETDTFLQPEGEV-----LVVADQPRDESTL-----QGINLGATFTVGYRSRLSH 137
Db 309 DSSSAFISQGFMAFPFGLSLTNMTDQSGNEANMVCSIQKPSRTSPIAFTTEENDYIVLNTS 368
Qy 138 FS-----RLEY-----WQPTTKGTDFWLYSPDQV-----HLGKSPQARISNPSQTT 182
Db 369 FSTFLVNCIDTGHQPVWQ-----LLAYSSPLILERSHLSETPOL-----411
Qy 183 QTAQWLLLEASVSSRGEQIYYQYRAE---DDTGCEADEITHLQATAQRYLHIVYGNRTA 239
Db 412 ----YYKYQVAPKPEDIFTNIEADLRADPSWLMQDQISLQLNRTATTFTSTLQIQYSSDA 467
Qy 240 SETLPGLDGSAPSQADWLFYLVFDYGERSNNLK---TPPAFTSTG-----SW 283
Db 468 QITLPRAE-MRPFVKHWTMI-----SRDNNTKLEHTVLVGGTVGLNCPGQGDPTPHVDW 520
Qy 284 LCRDQRFSEYGEF-----IRTRRLCRQVL-----MYHLQALDSKITEHNGPTLVS 331
Db 521 LLAGSKVRAPYVSEDEGRILDKSKGLQWADSFDTGVYHCIS-----565
Qy 332 RLILNYDESAIATSLVFRVRVGEQGNVVTLPPLLEYAQDFSPRRHAHW---QPMDEVLAN 389
Db 566 ----NYDDADILTY-----RITVVEPLVEAYQB-NGIHTVFIGETLIDLPC 607
Qy 390 FNAIQ-----RWQLVDLKGELPG---LLYQDKGAWWYRSQAQLGEIGSDAVTWKXQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKVLYNN-----GTLRILOV 646
Qy 444 IPSLQSNASLVYDINGDGLDGVITVTPGLRGVHSORP---DGSWTRFTPLNALPVEYTHPR 500
Db 647 TPKQGGYTRCVAAAPSG-VDFLIQVSVK-MKGQPLEHDE-----TEGSGIDESNPI 698
Qy 501 AQLADLMGAGLSDLVLI-----GPKSVELYANTRDGP---532
Db 699 AHLKEPPGAQLRTSALMAEYGVKHTSTSKRHNRYRELTLQRGDSVTHRRFRNRHPPPS 758

```

```

Qy 533 -----AKGKDVQSGDITLVPVGADPRKLVAFSDVLG-----SQOHLIVE 572
Db 759 ARRIDPOHWAALLEKAKKNAMPDKRENTVSP---PPVTQLPNIPGREDSSGMLALHE 815
Qy 573 ---VSATKVTCPNL-----GRGRFGQPIITLPGFSQPATFBNPAQVYLADLDGSGPTDL- 623
Db 816 EFWVPATKALNLPARTVADSRITSDSPMTNINY---GTESPV-VNSQILPPEPTDFK 871
Qy 624 --IYVHTNRLDIFLANKSGNGFAEPVTLR-----PP-----EGLRFDH----- 658
Db 872 LSTAIKTTAMSKNINPTMSSQIQCTTNQHSSTVFPLLLGATEFQDSDQMGCRGHEHFSRP 931
Qy 659 --TCOLOMADVQGLGVAS-----LILSVPHMSPHHWRCDLTNMPKWLNNMNNNGVHHT 711
Db 932 PITVRTWKVDNVKMLSTNKLLESVNTNSHQ-----TSVRE--VSEPRHNFYSHT 984
Qy 712 LRVSSSQFWLDEKAAALTTGQTPVCYLPPPIHTLWQTEDEISGNKLVTLRYARGAM 771
Db 985 TQILSTSTFPSPDPTAA-----HSQFP-----IPRSTVNIPLFRR--- 1020
Qy 772 DGREREPFGYVEQTDSHQLAQGNAPERTPPALTKWYATGLPVIDNALSTEWDRDQA 831
Db 1021 FGRQKIGRGRGI-----ISPYRT-PVLRHRYSI-----FRSTTRGSSEKS 1061
Qy 832 PAGES-----PRFTTWQ-----DNKOVPLTPEDDNSRYWFNRA 864
Db 1062 TTAFSATVNLVNTCULSCPRLRLITATAALSPPSAPIFFPKADIARVPSEST----- 1114
Qy 865 LKGQLRSELYGLDD--STNKHVEYVTEFRSQVRRLOHTDSRYFVLWSSVVESR---N 918
Db 1115 ---TLVQNPLLLLENKPSVEKTIPT-TIKYFRTEISQVTPGTGAVMTYAPTSIPMEKTHKVN 1170
Qy 919 YHYERIASDPQCSQNTLSSDRFG-----QPLKQLSV-QYPRRQOPAINLIPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSSLSGAIKTPMTIIAITFRSRRKIPW-----QQNFV 1221
Qy 974 NSYDDQOR---QLRLTYQSS-----WHHLTNNTVRVLGLP----- 1006
Db 1222 NNNPKRLRNQHKVSLQKSTAVMLPKTSPALPORQSLPSHHTTKTNPGSLTKKELP 1281
Qy 1007 -----DSTRSDIFTYGA-----ENVPAGGLNLELLSDKNSLIADDKPREY 1046
Db 1282 FPPLNPLMPSLIISKDSSTKSIISTQTAIPATTPFPFASVITYETQTSRSRAQTIQREQ- 1340
Qy 1047 LGOQKATYDQNTPLQ-----TPTRQA--LIAFTETTVNQSTLSAFNGSIIPSDK--L 1097
Db 1341 --POKQRTD-PNISPQSSGFTTPTAMTPPVLTAETSV--KPSVSAFTHSPPEPTTGI 1395
Qy 1098 STTL 1101
Db 1396 SSTI 1399

RESULT 15
US-10-453-372-734
; Sequence 734, Application US/10453372
; Publication No. US2006000323A1
; GENERAL INFORMATION:
; APPLICANT: Alsobrook, et al.
; TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
; FILE REFERENCE: 21402-589 A
; CURRENT APPLICATION NUMBER: US/10/453,372
; CURRENT FILING DATE: 2003-06-03
; PRIOR APPLICATION NUMBER: 09/789390
; PRIOR FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: 60/185967
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/823187
; PRIOR FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195792
; PRIOR FILING DATE: 2000-03-10
; PRIOR APPLICATION NUMBER: 09/839446

```

```
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/199476
; PRIOR FILING DATE: 2000-03-25
; PRIOR APPLICATION NUMBER: 09/863776
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: 60/208263
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/939398
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: 60/227800
; PRIOR FILING DATE: 2000-08-25
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1609
; SOFTWARE: CuraSeqList version 0.1
; SEQ ID NO 734
; LENGTH: 2617
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-453-372-734

Query Match
Best Local Similarity 1.6%; Score 127; DB 6; Length 2617;
Matches 242; Conservative 170; Mismatches 426; Indels 426; Gaps 67;

Qy 91 DETDTFLGPEGEV-----LVADQPRDESTL-----QGINLGATFTVTGYSRRLSH 137
Db 309 DSSAFISFQFMAPPFSLTLMNTQSGNEAMVCSIQKPSRTSPIAFTTEENDYIVLNTS 368
Qy 138 FS-----RLEY-----WPKTTGKTDFWLYSPDQV-----HLGKSPQARINPSQTT 182
Db 369 FSTFLVNCIDYGHIOFVWQ-----ILALYSDSPLILERSHLLSETPQL----- 411
Qy 183 QTAOWLLBASVSSRGEQIYYQVRAB---DDTCEADEITHLQATAQRYLHIVVYGNRTA 239
Db 412 ----YYKYQVAPKPEDIFTNEADLRADPSWLMQDQISLQLNRTATFTSTLIQYSSDA 467
Qy 240 SETPLGLDGSAPSQADWLFYLVDFYDGERSNLJK---TPPAFSTTG-----SW 283
Db 468 QITLPRAB-MRPVKHKWTMI-----SRDNTTKLEHTVLVGVTVGLNCPGQGDPTPHVDW 520
Qy 284 LCRQDRFSRYEGFE-----IRTRLCQVL-----MYHLQALDSKITEHNGPTLVS 331
Db 521 LLDGSKVRAPVYSEDGRILIDKSKLEQWADSPDTGVYHCIS----- 565
Qy 332 RLILNYDESATLSTLVFVRVGHQDGNVVTLPPLLELAYQDFSPRRHAHW--QPMVDVLAN 389
Db 566 ----NYDDADILTY-----RITVVEPLVEAYQB-NGIHTVTFIGETLDFCH 607
Qy 390 FNAIQ-----RWQLVDLKGEGLPG--LLYODKGAMWYRSAQRIGEIGSDAVTWKMQPLSV 443
Db 608 STGIPDASISWV-----IPGNVLYQSS-----RDKKVLNN-----GTLRILQV 646
Qy 444 IPSLQSNASLYDINGDQGLDWITGFLRGVHSORP---DGSWTFPTPLNALPVEYTHPR 500
Db 647 TPKDQGYTRCVAANPSG-VDFLIQVSVK-MKGQRPLEHDE-----TEGSGLDSENP 698
Qy 501 AOLADLMGAGLSDLVLI-----GPKSVRLYANTRDGF--- 532
Db 699 AHLKEPPGAQLRTSALMAEAVGKHT'SSTSKRNYBELTLQRGDSHRRFRNRHFFPS 758
Qy 533 -----AKGQDVQSGDITLVPVGDAPRKLAVFSDVLG-----SQQAHIVE 572
Db 759 ARRIPDQHWAAALLEKAKKNAMPDKRENTVSP---PPVVTQLPNIPGEEDDSSGMLALHE 815
Qy 573 ---VSATKVTWPNL-----GGRFGQITLPGFSQPATEFNPAQVYLADLDGSOPTDL- 623
Db 816 EFMVPATKALNLPARTVTADSTISDSPMTNINY---GTEFSPV-VNSQILPPEEPTDFK 871
Qy 624 ---IYVHTNRLDIFLNKSGNGRAEPVTLR-----FP-----EGLRFDH----- 658
Db 872 LSTAIKTAMSKNIINPTWSSQIQGTNTQHSVTFLLIGATEFQDSQDQMGREHFQSRP 931
Qy 659 ---TCOLOMADVQGLGVAS-----LILSVPHMSPHHWRCDLTNMKFWLLNEMNNNNVGHHT 711
```

```
Db 932 PITVTRTWIKDVNVKMLSTTNKLLSVNTTNSHQ-----TSVRE--VSEPRHHFYSH 984
Qy 712 LRYRSSQFMDDEKAAALTTGQTPVCYLPFPFIHTLMQTEDEISGNKLVTTLYARGAW 771
Db 985 TQILSTSTFPSPDPTAA-----HSQFP-----IPRNSTVNIPLFR--- 1020
Qy 772 DGRERFRFGYVBCQTDHSHQLAQNABERTPPALTUKWYATGLVIDNALSTETWRDQA 831
Db 1021 FGRQKIGGRGRI-----ISPYRT-PVLRHRHYSI-----FRSTTRGSSEKS 1061
Qy 832 FAGES-----PRFTTWO-----DNKDVLPTPEDDNSRYWFNRA 864
Db 1062 TTAFSATVNLVNTCULSPRERLTTATAALSPSAPITFPKADIARVSEEST----- 1114
Qy 865 LKGQLRSLEYGLDD--STNKHVPYVTFEFSQVRRLOHTDSRYPLVWSSVSES---N 918
Db 1115 ---TLVQNPLLELLENKPSVEKTP-TIKYFRTEISQVTPGTGAVMTYAPTIPMEKTHKN 1170
Qy 919 YHYERIASDPQCSQNIILSSDRFG-----QPLKQLSV-QYPRQOPAINLYPDTLPDKLLA 973
Db 1171 ASYPRVSTNEAKRDSVITSSLSGAIKPKPMTIITATFRSRKIPW-----QQNFV 1221
Qy 974 NSYDDQOR---QLRLTYQOSS-----WHHLTNNTVTVLGLP----- 1006
Db 1222 NNHNPGRRLRNQHKVSLQKSTAVMLPKTSPALPORQSLPSHHTTKTHNPGSLPTKKELP 1281
Qy 1007 -----DSTRSDIFTYGA-----ENVPAGLNLLELSDKNSLIADDPKPREY 1046
Db 1282 FPLNPMPLPSIISKDSSTKSIISTQTAIPATTPFPASVITVETQERSRAQTIQREQE- 1340
Qy 1047 LGQOKTAYTDGQNTTPIQ-----TPTROA--LIAFTETTVNQSTLSAFNGSIPSDK--L 1097
Db 1341 --POKKNRTD-PNISPQSSGFTTPTAMTPPVLTAAETSV--KPSVSAFTHSPPEPNTTGI 1395
Qy 1098 STTL 1101
Db 1396 SSTI 1399
```

Search completed: February 16, 2006, 21:55:16
Job time : 25.4867 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:31:49 ; Search time 21.432 Seconds
(without alignments)
4309.818 Million cell updates/sec

Title: US-10-754-115-47
Perfect score: 5005
Sequence: 1 MNIDPKLYQKTPVSVYDN.....DAEISFLTPIKVKVKEHR 960

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2169	43.3	874	2 AE0281	probable insectici
2	2138	42.7	984	2 AE0290	insecticidal toxin
3	1833.5	36.6	952	2 AC0447	probable insectici
4	1831.5	36.6	1011	2 AD0447	probable insectici
5	308	6.2	2334	2 S32920	cell wall-associat
6	307.5	6.1	843	2 AB0539	Rhs-family protein
7	304	6.1	1404	2 E85509	hypothetical prote
8	300.5	6.0	1426	2 H64780	rhgD protein precu
9	300.5	6.0	2183	2 T37218	hypothetical prote
10	300	6.0	1404	2 E90658	RhsG core protein
11	287	5.7	1438	2 AI0093	conserved hypot het
12	285	5.7	1409	2 F91187	rhaA core protein
13	283	5.7	1397	2 C64805	rhac protein precu
14	282.5	5.6	2167	2 AF1489	cell wall-associat
15	279.5	5.6	1512	2 AH0439	probable membrane
16	278.5	5.6	1411	2 E85145	rhac protein precu
17	277	5.5	1377	2 C85159	rhac protein precu
18	274	5.5	1397	2 A95570	rhac protein in rh
19	272.5	5.4	1399	2 A95720	RhsC core protein
20	272	5.4	1317	2 H83310	conserved hypot het
21	270.5	5.4	1354	2 AG0538	Rhs-family protein
22	270	5.4	1400	2 E90886	RhsG core protein
23	267.5	5.3	1394	2 H91236	RhsH core protein
24	261.5	5.2	985	2 B86084	hypothetical prote
25	258	5.2	1398	2 B85549	hypothetical prote
26	257.5	5.1	1377	2 E86034	rhac protein in rh
27	257	5.1	1398	2 H90698	RhsD core protein
28	247.5	4.9	1616	2 E90704	Rhs core protein w
29	242	4.8	1645	2 H85554	hypothetical prote

30	224	4.5	656	2 G85731	Rhs element associ
31	223	4.5	682	2 C84898	rhac protein - Rsc
32	221.5	4.4	456	2 AB0439	conserved hypot het
33	211	4.2	586	2 B90659	Rhs core protein (
34	211	4.2	586	2 H85509	hypothetical prote
35	211	4.2	794	2 T36972	probable membrane
36	180.5	3.6	1630	2 A53577	ascites sialoglyco
37	180	3.6	2468	2 AB3412	hypothetical prote
38	177	3.5	356	2 T37136	hypothetical prote
39	176	3.5	1959	2 AG1085	hypothetical prote
40	174.5	3.5	1417	2 H90670	probable invasiv e
41	174.5	3.5	1417	2 D85521	probable adhesin e
42	174	3.5	4688	2 F82885	hypothetical prote
43	173.5	3.5	709	2 S38241	hypothetical prote
44	172.5	3.4	901	2 A49227	sialidase - Actino
45	172	3.4	4936	2 AH2515	hypothetical prote

ALIGNMENTS

RESULT 1

AI0281
probable insecticidal toxin complex YPO2312 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AI0281
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AI0281
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-874 <R>
A:Cross-references: UNIPROT:Q8ZE71; UNIPARC:UPI00000DC6A3; GB:AL590842; PIDN:CAC91117.1
C:Genetics:
A:Gene: YPO2312

Query Match	43.3%	Score	2169;	DB 2;	Length	874;			
Best Local Similarity	53.0%;	Pred. No.	3.3e-115;						
Matches	451;	Conservative	106;	Mismatches	202;	Indels	92;	Gaps	15;
Qy	8	LYQKTPVSVYDNRGLIIRNIDFHRHTANGDPDT-----RTRHQYDIHGLHMQSIDPR	61						
Db	5	LFSKTPSVTVLNRGLTVRDIAYHR-----HPDSPDVISERTHHQYDARGFLTQSADPR	59						
Qy	62	LYEAKQTNNTIKPNFLWQYDLTGPNPLCTESIDAGRTVTLNDIEGRPLLTVT-----	112						
Db	60	LHGAGLM-----NFSYLTDLTGRIILRTQGADNGTTVSLNDAAGRPFISASNISTSDGT	113						
Qy	113	---ATGVITQROYETSSLPGRLLSVAEQTPPEKTSRITERLLIWAQNTAEKHNLAGQCV	169						
Db	114	EDRQAMRTWQYEASLPGRLLSVTEQV-TGKATRIYERFYAANTDAEKSINLAGACV	172						
Qy	170	RHYDTAGVTRLESLSLTGTVLSQSSQLIDTQE-----ANWTDGNETVWQNMUADDIYTL	225						
Db	173	SHYDTAGLVQPSIALTCVPLSVTRRLMKSADNPDAVDWQWQADASAWNQLDGTHTSL	232						
Qy	226	STFDATGALLTQTDAGNIGQRLAYDVAGQLNGSWTLKGQTEQVILKSLTYSAAGQKLR	285						
Db	233	TTADATGAVLTATTDAGNIGQRLAYDVAGQLNGSWTLKGQTEQVIVKSLTYSAAGQKLR	292						
Qy	286	EHGNDVITEYSEVPEPTORLIGIKTRP-----SDTKVLQDLRYEYDVPVGNVIRINDAET	341						
Db	293	EHGNGVVTYYEYEPETQRLVGIKTERPAGHASGNKVLQDLRYEYDVPVGNVLKVTINDAET	352						
Qy	342	RFWNNQKVPENTYTYDSLYQLISATGREMANIGQSHQFPSPA--LPSDNNYTYNTRYT	399						
Db	353	RFWRNQKVPENTYTYDSLYQLYSATGREMANAGQSCSPSTTVPPLPAPDASATRYISRT	412						

QY 400 YTDGNGNLTKI QHSSPATONNYTNTITVSNRSNRAVLSTLTEDPAQVDALFDAGGHQNT 459
 DB 413 YTYDEAGNLTOIRNAPATNNSYTKITVDSNRGVLSTLUTENHADVDALFTAGGQOTK 472
 QY 460 LISQNLNWNTRGELQOVTLVKRDGANDREWYRSGDGRMLKINEQQAASNNAAQTV 519
 DB 473 LOPQHLLIWTARNELLKVTVPVRD-GSTDDSESYRYDAASQRIKLVSRQKNTSMQTV 531
 QY 520 TYLNLRLRLTONSTATTEDLQVITVGEAGRAQVRLVHWSGKPEDIDNNOLRYSDNLI 579
 DB 532 LYLPGELRLSKSGDTEGLQVITVGEAGRAQVRLVHWSGKPEDIDDDQIRYSYDNL 591
 QY 580 GSSOLELDSGOIISSEBEYYPYGGTALWAARNOQTEASYKTIYRSGKERDAGLTYGYRY 639
 DB 592 GSCLELGGDNIISABEYYPYGGTAVWARRAVEADYKTVYRSGKERDAGLTYGYRY 651
 QY 640 YQPWIGRWLSDPAGTIDGLNLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAP 699
 DB 652 YQPWAGRWLSADPAGWVDGLNLFWRARNPVAFIDRNGLSNELYS-QAFKRT----- 703
 QY 700 SPANATVAINIRPPVAPKPSLKASTSSQPTTHPIGAANIKP--TTSGSSIVAPLSPVG 757
 DB 704 --ANKYVNIIGVRAP-----NPLGETLLKEGFPKSNFHKAKSSPTG 743
 QY 758 -----NKSTSEISLPESAQSSSSTSTNLQKSFYLRADNRSFE----- 798
 DB 744 PTAGFIABDPYKVSFSAKQKASIDKAKALGESIDLFISKRNELIDTGNLNSLG 803
 QY 799 --EMQSKFPPEG 807
 DB 804 ENRYSAPKPYG 814
 RESULT 2
 AE0290
 insecticidal toxin [imported] - Yersinia pestis (strain CO92)
 C;Species: Yersinia pestis
 C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C;Accession: AE0290
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A;Reference number: AB0001; MUID:21470413; PMID:11586360
 A;Accession: AE0290
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-984 <KUR>
 A;Cross-references: UNIPROT:Q8ZE10; UNIPARC:UPI00000CDBFE; GB:AL590842; PIDN:CAC91185.1;
 C;Genetics:
 A;Gene: sepc
 Query Match 42.7%; Score 2138; DB 2; Length 984;
 Best Local Similarity 47.7%; Pred. No. 2.3e-113;
 Matches 473; Conservative 137; Mismatches 272; Indels 110; Gaps 24;
 QY 4 IDPKLYQKTPVSYVNRGLIIRNIDFHTTTANGDPDT--RITRHQYDIHGLNQSIDPR 61
 DB 1 MNTSLFSKTPAVTLNDRGLSVRSIAVHRHPSDP-DTGERITHQYDARGFLQSDAPR 59
 QY 62 LYEAQNTNTIKPNFLWQYDITGNPLCTESIDAGRTVTLNDIEGRPLLTWT----- 112
 DB 60 LHD-----TGRANVKYLSDLVGNALCTVSADVGTVALNDAAAGLFEFTVSNIDTADDDL 113
 QY 113 ---ATGVIQTYQYETSSLPGRLLSVAEQTPREKTSRITERLIWAGNTAEAKDHLAGOCV 169
 DB 114 EDRSQVNRRTWQYEGASLPGRLLSITEQVTGE--AARVTERTYAANTDEKALNLVGCQV 172
 QY 170 RHYDTAGVTRLESLSLTGTVLSQSSQLIIDTQE-----ANWTGDNETVWQNMLADDIYTL 225
 DB 173 SHYDTAGLRQMDSTALTGVLPSVTRRLIKDADNPDTVANMQGEDSSAWNDQLAADRLTTL 232

QY 226 STFDATGALLTOTDAKGNIOKLAYDVAGQNLGSMILTKGOTEQVVIKSLTYSAAGQKLR 285
 DB 233 TTADAI GAVLTITTKGNVRQVYVNVAGLLSGSWLRVKGGAEOVIVQSLTYSAAGQKLR 292
 QY 286 EHGNVDITEYSYEPETQRLIGIKTRR----PSDTKVLQDLRYEYDPVGNVJISIRNDABAT 341
 DB 293 EHGNVVTVTYBETQRLTGIRTERSGASHGSKVLQDLRYEYDPVGNVQSIIRNDABET 352
 QY 342 RFWNQKVMENVTYDLSQLISATGREMANIQOOSHOPSPALPSPDNNTYNTYTRYT 401
 DB 353 RFWRNQKVPENTYVYDLSYQLVSATGRVMVNAQOGRSLPSATLPLESSAYTNTYTHYT 412
 QY 402 YDRGNLTQIHSQSPATONNYTNTITVSNRSNRAVLSTLTEDPAQVDALFDAGGHQNTLI 461
 DB 413 YDTAGNLTOIRH-TPATGSGHTTDTITVDSNRGVLSTLTNPAEVDALFTAGGQKQLQ 471
 QY 462 SGQNLNWNTRGELQOVTLVKRDGANDREWYRSGDGRMLKINEQQAASNNAAQTV 521
 DB 472 PGQHLIWTARNELLKVTVPVRD-GSDDDSESYRYDGNRSQRIKLVSVQKGTGSTQTV 530
 QY 522 LPNLRLRLTONSTATTEDLQVITVGEAGRA--QVRLVHWSGKPEDIDNNOLRYSDNLI 579
 DB 531 LPRLELRSTASGVVETTESLQIITVGEAGRAQVQVQVVLHWEKGPDAINDOLRYSDNLI 590
 QY 580 GSSOLELDSGOIISSEBEYYPYGGTALWAARNOQTEASYKTIYRSGKERDAGLTYGYRY 639
 DB 591 GSSTLEVDGDNVLSMEEYYPYGGTAVWARRAVEADYKTVYRSGKERDAGLTYGYRY 650
 QY 640 YQPWIGRWLSDPAGTIDGLNLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAP 699
 DB 651 YQPWGWRWLSADPVGTVGLNLYLMVGNPNPSTPHDSNGLI-----REGQSARKLVGEAFV 705
 QY 700 SPANAT-----NVAINIRPP-----VAPKPSLKASTSSQPTTHP----- 734
 DB 706 HPLEMSYFERISIBENMAMSVREAGIYITLALGEGAAAGHNIUKTIKPGSLKAVENK 765
 QY 735 IGAANIKPPTSG-----SSIVAPLSPVGNKSTSEISLPESAQ-SSSSSTSTNLQK 784
 DB 766 AGALELAKNSGFGTRVGRWNASGVQGVYVNRSGEDLVYPASLQDTSDELNVNAMIKH 825
 QY 785 KSFTLYRADNRSFEEMOSKPEGKAWTP-----LDTKMARQFASIFIGQKDTSNLPKETV 840
 DB 826 KIIFPYTGVDYDMHDI--KFNRG-KGVVPTAESAEETGVKDLINKGVAEVPARPFEYTA 882
 QY 841 KNISTWGAQKPLKDLNVIKTKDKSTVWVSTAINTEAG-----QSSGAPLHKIDM 892
 DB 883 MNVIRHG--PQVNFVPMWYEHDK-----VVSNGYLVGVVARPGPFPIAMVHQGW 932
 QY 893 DLVEFAIDGOKL-----NPLPEGRTKNMV 916
 DB 933 TVFD---DSKELFNFKSSNTPLFEHWQONFI 961
 RESULT 3
 AC0447
 Probable insecticidal toxin YPO3673 [imported] - Yersinia pestis (strain CO92)
 C;Species: Yersinia pestis
 C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C;Accession: AC0447
 R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
 Nature 413, 523-527, 2001
 A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A;Reference number: AB0001; MUID:21470413; PMID:11586360
 A;Accession: AC0447
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-952 <KUR>
 A;Cross-references: UNIPROT:Q8ZAV9; UNIPARC:UPI00000DC7E2; GB:AL590842; PIDN:CAC93143.1;
 C;Genetics:
 A;Gene: YPO3673

Query Match 36.6%; Score 1833.5; DB 2; Length 952;
 Best Local Similarity 44.1%; Pred. No. 4.2e-96;
 Matches 436; Conservative 138; Mismatches 294; Indels 121; Gaps 27;

QY 3 NIDPKLYQKTPVSVYDNRGLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSIDPRL 62
 DB 4 SLPTOLCANTPALTTHNRGLAIFRLAYNRDRDNETVDELISRNRYNASGGLIASRDPRL 63

QY 63 YEAKQTNNTIKPNFLWQYDLTGNCPLCTESIDAGRVTIINDIEGRPLLTATGVTQY 122
 DB 64 E-----VDNFYQYSLSGVPLRTDSVDSGSTLQADLSAGRTVLTLDHHTHRWVEY 114

QY 123 ETS--SLPGRLLSVAEQTPPEKTSRITERLIWAGTEAKHNLKAGQCVRHVDYAGVRL 180
 DB 115 ETGEHSL-GRPLSYHEQA-KGGLKTVTDTRFFVATNSEQDKNCLNGQCVRHYDSAGLQAL 172

QY 181 ELSLSITGVLSQSSOLLIDTQ-EANWTGDNVWQNLADDIYTLSTFDATGALLTQTD 239
 DB 173 ISQSIIGVPLQOORRELLNPKGPFVDFEKEN-WGARLSEQPFVSHSTTDALGQLLTQTD 231

QY 240 AKGNITQRLAYDVAGOLNGSWLTKGTEQVILKSLTYSAAQOKLREHGNVDVITEYSYEP 299
 DB 232 AKGHIQRMAYNRAGOLIGSWLTKNSAEQVILRSITYSAAQOKLREESGNGVITEYRVEP 291

QY 300 ETQRLIGIKTRRPS----DTKVLQDLRYEYDPVGNVISIRNDAEATRFVHNOKVPENTY 355
 DB 292 QTQRLIGIKTRPAKDRPTR-LQDLRYDYDPVGNILAHNDAEATRFVHNOKIVPETTY 350

QY 356 TYDSLYQLISATGREMANIQOSHQPSPALPSDNNTYNTYTRTYDGRGNLTQIHS 415
 DB 351 RYDALYQLIEATGREATNGIQNSQLPALASLNSNQFVNYTRSYHYDRAGNLLKIQHTG 410

QY 416 PATONNYTNTIVSNRNRVSLTLEDPAQVDALPDAGGHONTLISGONLWNTRGELQ 475
 DB 411 ---ASQYSTHITVSDSSHGIIQOQGIITARDIRSQFDAGNQOQLOPQPLRNSNQLQ 467

QY 476 QVTLVKRDKGANDREWRYSGDGRMLKINEQOASNNAAQTORVYLPNLELRLQNSTA 535
 DB 468 QVEPVRNDGISDS-ESVLYDGGGRVAKISLHKTHNAIQTRSVIYLAGLELRGQNDNN 526

QY 536 TTEDLOVITVGBAGRAQVRVLHWESGKPEDIDNNQIRYSYDNLIGSSQLELSEGOIISE 595
 DB 527 LTESPQVITVGAAGRAQVRVLHWESGQPDVINDQRLYSFDFNHLGSALIELDSGDGIIISQ 586

QY 596 EBYEYVGGTALWAARNQTEASVKTIRYSGKRDATGLYYGYRYQYQPIGRWLSDDPAGT 655
 DB 587 EBYEYVGGTAVLASRNTVEAKYTVRYSGKRDATGLYYGYRYQYQPIGRWLSADPAGT 646

QY 656 IDGLNLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAPSANATNVAINIRPPV 715
 DB 647 IDGLNLYRMVRNPNIRWRDNGLL--TEBQI-----NMVNLFSNI 685

QY 716 A-----PKPSLPKASTSQPHTPIGAANIPTTSGSIVAPLSPVGNKSTSEISLPES 769
 DB 686 GLKNDDELKSELLKYGLSEB-PQNQIYLNMLRPMQSGS-----SSLSFPSE 732

QY 770 AQSSSSSTSTNLQKSFYLRADNRSPFEMQSKPEGFKAWT---PLDTKMARQFASIF 826
 DB 733 SSSSSSQSQVD-----SGVLSFVRNVH-----PFEDIKATMRPVPKQASDDITY 781

QY 827 IQOKDTSNLPKE-----TVKNISTWGAQPKLKDLSNYIKY-TKDK-----S 866
 DB 782 SAEDITEASPIKILIGDLTSTKNTQY--KSALAEGK-IKYITKEKEYEITDPFEGGLS 837

QY 867 TVWVSTAIN-----TEAGQSSGAPLHKIDMDIYEPFAIDGQKLNPLPEGR 912
 DB 838 TEQIDLTWNKILKQKOLVGHICGAGNRSQVIAALSIN-KQYTTD--KINSFDV--T 892

QY 913 KMWVPSLLDTPQIETSSIIALNHPQVND 941
 DB 893 HSLRGSILKDTQYQVDTVTAKAVGIIE 921

RESULT 4
 AD0447
 C:Species: Yersinia pestis
 C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
 C:Accession: AD0447
 R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
 deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,
 Nature 413, 523-527, 2001
 A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
 A:Reference number: AB0001; MUID:21470413; PMID:11586360
 A:Accession: AD0447
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1011 <KUR>
 A:Cross-references: UNIPROT:Q8ZAV8; UNIPARC:UPI00000CDA64; GB:AL590842; PIDN:CAC93144.1.
 C:Genetics:
 A:Gene: YPO3674

Query Match 36.6%; Score 1831.5; DB 2; Length 1011;
 Best Local Similarity 46.0%; Pred. No. 6e-96;
 Matches 422; Conservative 130; Mismatches 278; Indels 87; Gaps 23;

QY 1 MNIDP-KLYQKTPVSVYDNRGLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSID 59
 DB 69 MPNITLPTDLCANTPTLAIHNRGFAIRTLAYNRDRDNETIGELISRNRYNASGGLIASRD 128

QY 60 PRLYEAKQTNNTIKPNFLWQYDLTGNCPLCTESIDAGRVTIINDIEGRPLLTATGVTQ 119
 DB 129 PRLE-----VDNFYQYSLSGVPLRTDSVDSGSTLQADLSAGRTVLTLDHHTRW 179

QY 120 RQYETS--SLPGRLLSVAEQTPPEKTSRITERLIWAGTEAKHNLKAGQCVRHVDYAGV 177
 DB 180 VEYETGEHSL-GRPLSYHEQA-KGGLKTVTDTRFFVATNSEQDKNCLNGQCVRHYDSAGL 237

QY 178 TRLESLSITGVLSQSSOLLIDTQ-EANWTGDNVWQNLADDIYTLSTFDATGALLT 236
 DB 238 QALINQSIIGVPLQOORRELLNPKGPFVDFEKEN-WGARLSEQPFVSHSTTDALGQLT 296

QY 237 QTDKAGNITQRLAYDVAGOLNGSWLTKGTEQVILKSLTYSAAQOKLREHGNVDVITEYS 296
 DB 297 QTDKAGHIQRMAYNRAGOLIGSWLTKNSAEQVILRSITYSAAQOKLREESGNGVITEYR 356

QY 297 YEPETQRLIGIKTRRPS----DTKVLQDLRYEYDPVGNVISIRNDAEATRFVHNOKVP 352
 DB 357 YEPQTORLIGIKTRPAKDRPTR-LQDLRYDYDPVGNILAHNDAEATRFVHNOKIVPE 415

QY 353 NITYYDLSYQLISATGREMANIQOSHQPSPALPSDNNTYNTYTRTYDGRGNLTQI 412
 DB 416 TTYRYDALYQLIEATGREATNGIQNSQLPALASLNSNQFVNYTRSYHYDRAGNLLK 475

QY 413 HSSPATONNYTNTIVSNRNRVSLTLEDPAQVDALPDAGGHONTLISGONLWNTRG 472
 DB 476 HTG---ASQYSTHITVSDSSHGIIQOQGIITARDIRSQFDAGNQOQLOPQPLRNSN 532

QY 473 ELQOQVTLVKRDKGANDREWRYSGDGRMLKINEQOASNNAAQTORVYLPNLELRLQ 532
 DB 533 QLOQVEPVRNDGISDS-ESVLYDGGGRVAKISLHKTHNAIQTRSVIYLAGLELRQ 591

QY 533 STATTEDLOVITVGBAGRAQVRVLHWESGKPEDIDNNQIRYSYDNLIGSSQLELSEGOI 592
 DB 592 GNNLTDFQVITVGAAGRAQVRVLHWERGQPDVINDQRLYSFDFNHLGSALIELDSGD 651

QY 593 ISSEYYPYGGTALWAARNQTEASVKTIRYSGKRDATGLYYGYRYQYQPIGRWLSDDP 652
 DB 652 ISQSEYYPFGGTAVLASRNTVEAKYTVRYSGKRDATGLYYGYRYQYQPIGRWLSADP 711

QY 653 AGTIDGLNLYRMVRNPNVTLDPDGLM--PTIAERIAALKKNKVTDSAPSANATNVAIN 710
 DB 712 AGTIDGLNLYRMVRNPNVGLMDGDLMTDKLLAKHEANFPAKKNISSMAELKSEIEKGL- 770

QY 711 IRPPVAPKPSLPKAS-----TSSQPTHTPIGAANIPTTSGS-IVAPLSPVGNKS-- 760


```
QY 38 -----DPDTRITRHOYDIHGLNQSIDPRLYEAKQTNTTKPNFLWOYDLTGPNLCTES 91
Db 219 RLERTVDAESNSTRYSDGNHGLTEV-----MFADGRTER-----YQPDAGR-LVKYT 266
QY 92 IDAGR-TVTLNDIEGRPLLTATGVIQTRQYETSSLPGRLLSVAEQTP-----EKTS 144
Db 267 SPAGQITWRQDQGRVRROTDATGRTAYEYDAY---GRLLTLTNENGESYRFYDYVLD 323
QY 145 RITER-----LIWAGNTEAEKDHNLACQCVRHRYDTAGVTRLESLS 184
Db 324 RVTEQDPGSPRAYGNALNAVAVIYGGEGEIRHGL-----ERDAAG-----R 370
QY 185 LTGTVLSQ-----SSQLLIDTO-----EAWTGDNETVMQNMADDIYTLSTPDATG 232
Db 371 LTAKITPTREYRYDAADRLLIEIRRRHDAEGEPEVI-----RFSYDSAG 418
QY 233 ALLTQTDAGNIQRLAYDVAGOLGNSWLTGKQTEQVLIKSLITYSAAGKLRBEHNDVI 292
Db 419 NLLSEBTAQGVLOH-RYDVQGNRTETOMP-DGRTLYRL-----YVSGHLQOINLGRDVI 471
QY 293 TEYSYE-----PETQRLIG-IKTRRPD-----TKVL-----QD- 320
Db 472 SEPTFDHLHREVQSGRLDMRWYDRTGRLTRKLTCKMGVVPETPIDREYAYSGDE 531
QY 321 -----LRYEYDPVGNVISIRNDAEATRFMHNQKVPENTTYDSLYQLISATGR 369
Db 532 LLKKRHSRQGVTDYFYDTTGRITACRNEAYL-----DSWQYDAAANLLDRRQ 579
QY 370 EMANIGQOSHQPSPALPSDNTTYNTTFT-YTYDRGGLNLTQI QHSSPATQNTTYNTTV 428
Db 580 ETAQAGAGS-----VWPF--NRITSYRGLHYRYDEYGRV----- 612
QY 429 SNRSRAVLSTLTEDPAQVDALFDAGGHQNTLISQGNLNNWTRGELQOVLTKRDKGAND 488
Db 613 -----EKGRNGT-----QHYRWAEHRLTEVAVT---RGTV 642
QY 489 DREWYSGDGRMLKINEQASNAQTORVLYLPNLELRLTQNSTATTEDLQVITVGEA 548
Db 643 RRYGVYDAPGRVBEK-HELDAGKPY-NRTTFLWD-QWRLAQEC-----RL 686
QY 549 GRAQVRLHWSGKPEDI-----DNNQLRYSDNLISSQLELDSGQIIEEY 599
Db 687 GRSSLYTYDRSGSHEPLARVDRAAPGEADVLYYHTDVNGAPEMTDGGGNIWNEAGYQ 746
QY 600 PYGGTALWAARNQTEASVKTIRYSGKRD-ATGLYYGYRYQYQWIGRWLSSDPAGTIDG 658
Db 747 VAGNL---THEKETPVQONLRFQOYLDREGLHNLRYFYDPDIGHFISGDPGLAGG 803
QY 659 INLYRMVRNPNVTLDDPGLMPTIA 683
Db 804 INLYQYA-PNPLSYIDPLGLCKKFA 827

RESULT 7
85509 hypothetical protein 20268 [imported] - Escherichia coli (strain O157:H7, substrain EDL959)
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: E85509
R:Perna, N.J.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, B.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E85509
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1404 <STO>
A:Cross-references: UNIPROT:O8XED9; UNIPARC:UPI00000D025B; GB:AE005174; NID:g12512977; E
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
C:Gene: 20268
C:Superfamily: rhesF protein
```

Query Match 6.1%; Score 304; DB 2; Length 1404;
Best Local Similarity 21.0%; Pred. No. 5.4e-09;
Matches 192; Conservative 120; Mismatches 320; Indels 284; Gaps 44;

```
QY 18 YDNRG--LIIRNIDPHRTTANGDPDTRIT-----RHQYDIHGLNQSIDPRL--- 62
Db 646 YDAAGRITVLTNENGOSOTFRYDPVDRLTQORGDFGRTQRYHYDLTKLTQSEDEGLVT 705
QY 63 --YBAKQ--TNWTK--PNFLWOYDLTGPNLCTESIDAGRTVTILN---DIEGRPLLTVA 113
Db 706 WHYDASDRITHTVNGDPAEQWYDEHGLWTLTSHTEGHRVSVHYGYDDKGR-----L 759
QY 114 TGVITQRYETSLPGRLLSVAEQTPPEKTSRITERLIWAGNT-BAEKDHNL-----G 166
Db 760 TGERQT-----VENPE-----TGEMLWEHETGHAYSEQGLATRQEPDG 797
QY 167 QCVRHDTAGVTRLESLSLTGTVLSQSSQLLIDQEANWTDGNETVMQNMADDIYTLT 226
Db 798 LPPVWLTYSGGYLAGMKLGCTPLVEYMRDLRHRETARSFG-----GEAYELAT 846
QY 227 TFDATGALLTQTDAGNIQRLAYDVAGOLGNSWLTGKQTEQVLIKSLITYSAAGKLRBE 286
Db 847 AWNTSGQLRSR--HLNLPQLDRYDWDNDGQLIRISGPQES----- 885
QY 287 HGNDVITEYSYEPETQRLIGIKTRRPDVKYLODLRYEYDPVGNVI---SIRNDAEATRF 343
Db 886 -----REYRYS-DTGRLTGVHT---TAANLDIDIPYATDPAGNRLPDPPLHDPSTLTA- 934
QY 344 WHNQKVPENTTYDSLYQLISATGREMANIGQSHQFPSPALPSDNTTYNTTYTYTD 403
Db 935 WPDRIAEADAHYV--RYD-----EYGLAERTDRIPEGVI---RMHDERTHHHYD 981
QY 404 RGGNL---TKIOHSPATQNTTYNTITVSNRSRAVLSTLTEDPAQVDALFDAGGHQNTL 460
Db 982 SQHRLVFHTRLOHSEPQVESRY-----LVDPLGR--- 1011
QY 461 ISGNLNNWTRGELQOVLTKRDKGANDREWYSGDGRMLKINEQASNAQTORV 520
Db 1012 -TGKRV-WRRERDLTGWMSLSR---KPEETWYGDGD--RLTTVQTQTRIQTIVYQPGS 1063
QY 521 YLPNLELRLTQNSTATTEDLQVITVGEAGRAQVR---VLHWESG----- 561
Db 1064 FTPLLRIE-TEN-----GEQAKARHSLAEVLQEDTGVTLPABELAVMLGRLE 1109
QY 562 ---KPEDIDNNQLRYSDNLISSQLELDSGQIIEEY-----PYGG 603
Db 1110 RELRQGSVSEESQWLACQGLTAEQMAAQLEAEYIPERKHLHYCHDRGLPLALISPEGE 1169
QY 604 TALWAAR-----NOTEASY--KTIRYSGKRD-ATGLYYGYRYQYQWIGRWLSSDP 652
Db 1170 TA-WQGEYDEWGNLLGETSAHQLOQSLRLPQQYDEESGLYNNRNYRYPDPLQGRYITQDP 1228
QY 653 AGTIDGLNLYRMVRNPNVTLDDPGLMPTIAERTAALKKNKVTDSAPSNATVAINR 712
Db 1229 IGLEGGNLLYOYPL-NPIEHIDPLGL-----ALDNLNYSDDPIYKGLSLN----- 1273
QY 713 PPVAPKPSLPKASTSSOPTHPIGAANIKPPTSGSSIVAPLSPVGNKSTSIPLSPESAQS 772
Db 1274 -----REPTGFTVGGHG-SPTSMDDRI-----KKGSDLTIKQLASD 1310
QY 773 SSSS-----TTSTNLQKGSFTLYADNRSP-----EENQSKPPGPGFKAWTPLD 815
Db 1311 IRANPKYHEGMPVVLFCSETGKGKNSPAQKLANELDATVIAPDEIITWIPDGNVA----- 1365
QY 816 TKMARQFASIFIGKD 831
Db 1366 --IMGOTARITIGKD 1379
```

RESULT 8

H64780
rhesD protein precursor - Escherichia coli (strain K-12)

C;Species: Escherichia coli
C;Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 09-Jul-2004
C;Accession: H64780; J50625; B30092; I69401; S16026
R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A;Title: The complete genome sequence of Escherichia coli K-12.
A;Reference number: A64720; MUID:97426617; PMID:9278503
A;Accession: H64780
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1426 <BLAT>
A;Cross-references: UNIPROT:P16919; UNIPARC:UPI0000047C9A; GB:AE000156; GB:U00096; NID:9278503
A;Experimental source: strain K-12, substrain MG1655
R;Sadosky, A.B.; Gray, J.A.; Hill, C.W.
Nucleic Acids Res. 19, 7177-7183, 1991
A;Title: The Rhd-E subfamily of Escherichia coli K-12.
A;Reference number: J50625; MUID:92115567; PMID:1766878
A;Accession: J50625
A;Status: nucleic acid sequence not shown
A;Molecule type: DNA
A;Residues: 1-100 <SAD>
A;Cross-references: UNIPARC:UPI000016F486; GB:M21764; GB:J04224; NID:9147646; PIDN:AAA2424
R;Feulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.; J. Bacteriol. 172, 446-456, 1990
A;Title: Structure of the rhaA locus from Escherichia coli K-12 and comparison of rhaA with other rhaA genes.
A;Reference number: I54935; MUID:90094253; PMID:2403547
A;Accession: I69401
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1232-1426 <RSS>
A;Cross-references: UNIPARC:UPI000000047B; GB:M29719; NID:9147644; PIDN:AAA24541.1; PID:9278503
C;Comment: the rha core consist of two distinct parts: a large N-terminal core that is conserved in all rha genes and a small C-terminal core that is conserved in all rha genes.
C;Genetics:
A;Gene: rhd
C;Superfamily: rhp protein
C;Keywords: transmembrane protein
F;1-26/Domain: signal sequence #status predicted <SIG>
F;27-1426/Product: rhaD protein #status predicted <MAT>
F;28-55/Domain: transmembrane #status predicted <TM>

Query Match 6.0%; Score 300.5; DB 2; Length 1426;
Best Local Similarity 21.7%; Pred. No. 8.8e-09;
Matches 186; Conservative 101; Mismatches 294; Indels 275; Gaps 42;

Qy 32 RTTANGDPDTRITRHOYDIGHNLQSIDPRL-----YEAKQ--TWNTI--KPNFLWQYDL 82
Db 667 RLVOQGGFGRTQRTHTGLTQSEDEGLVILWYDESRITRTVNGEPAEQWQIDG 726

Qy 83 TGNPLCTESIDAGRTVTLLN--DIEGRPLLTATGVIQTRQYETSLPGRLLSVAEQTP 139
Db 727 HGWLTDIHLSEGRHVAHYGYDDKGR-----LTGEQQT-----VENP 764

Qy 140 EKTSTRITERLIWAGNTE-AEKDHLNAGCQVR-----HYTAGVTRLESLSLCTVLSQ 192
Db 765 E-----TCELLWQHETKHAYNEQGLANRVTPLPPVLEWLTGSGYLAGMKLGCTPL-- 816

Qy 193 SSQLLIDTQEAQNTGCD--NETV--WQNLADD-IYTLTSTFDATGALLTQTDAGKNQTR 246
Db 817 -----VEYTRDLRHRTVRSFGWAGNAAEYLTSTYTPAGLQSQ-----HLNS 861

Qy 247 LAY--DVAGQLNSWLTGQTEQVILKSLTYSAAGQKLREHGNNDVITEYSYEPETQRL 304
Db 862 LVYDRDYGWSNDGLVRISGPRQ-----TREYGYG-ATQRL 896

Qy 305 IGKTRPSDKVLQDLRYEY--DPVGNVI---SIRNDAEATRFWQKQVMPENTYTYDS 359
Db 897 ESVRTLAPD-----LDRIYATDPAGNRLPDPPELHPDSTLT--VPDNRIADAHYV-- 948

Qy 360 LYQLISATGEMANIGQSHQSPFALPSPDNNTYNTYTRTYDRGNNL-----TKIQHSSP 416
Db 949 -----RHDEYGRLTETKDRIPAGVIRTDDE-----THHYHDSQHLRVFTVRIHQGE 997

Qy 417 ATQNNYNTITVNSRNVAV-----LSLTEDPAQVD 448
Db 998 LVESRYLYD--PLGRMAKRVWRERDLTGWMSLSRKEPEVTWYWGMDGRLTQTDTTRIQ 1056

Qy 449 ALPDAGGHQNTLISQNLWNTRGELQOVTLVKDKGANDREWVRYSGDGRMLKINEQ 508
Db 1057 TVYEPG-----SFTPLIIRVETENGERE-----KAQRSLAETLQ 1091

Qy 509 QASNNAAQ-----TORVYLPNLELRLTONSTATTEDLQVITVGBAGRAQVRVLMH--ESG 561
Db 1092 EGSNGHGVVPAELVRLDLE-----EIRADRVSSESA-----WLAQCG 1134

Qy 562 KPEDIINNQLRYSDNLISSQLDSEG-----QIISSEYYPYGGTALWAAR-----NOT 613
Db 1135 LTVEQLARQVEPEYTPARKAHLXCHDRGLPLALISED-----GNTAWSAEYDEWGNOL 1189

Qy 614 EAS-----YKTIRYSGKERD-ATGLYVYGVRYYOPWIGRWLSSDPAGTIDGLNLYRMVNR 667
Db 1189 NEENPHVYQYRULPGQOHDEESGLYNNRHHYDPLQGRYITQDPMGLKGNWLYQYPL- 1247

Qy 668 NPVTLLDPDGLMPTIABERIALAKKNKVTDSSAPSPANATNVAINTIRPPVAPKPSLPKASTS 727
Db 1248 NPLQOIDPMGLLQI-----WDDARSGACTGGVC----- 1275

Qy 728 SQPTHTPIGAANIPTTSGSIVAPLSPVGNKSTSEISLPESAQSS-----SS 775
Db 1276 -----GVLRSIIIGPSKFDSTADALDAL--KETQNRSLCNDMEYSGIVCKDTNGKYFAS 1327

Qy 776 STTSTNLQKKSFTLYR 791
Db 1328 KAETDNLKESYPLKR 1343

RESULT 9
T37218
hypochemical protein SC2H4.02 - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text_change 09-Jul-2004
C;Accession: T37218
R;Oliver, K.; Harris, D.; Parkhill, J.; Barrall, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1998
A;Reference number: Z21615
A;Accession: T37218
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-2183 <OLI>
A;Cross-references: UNIPROT:O86585; UNIPARC:UPI000000DADB3; EMBL:AL031514; PIDN:CAA20596.
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC2H4.02

Query Match 6.0%; Score 300.5; DB 2; Length 2183;
Best Local Similarity 23.0%; Pred. No. 1.8e-08;
Matches 198; Conservative 112; Mismatches 380; Indels 171; Gaps 40;

Qy 17 VYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIGHNLQSIDPRLYEAKQNTNITKFNF 76
Db 1267 VTDAFGRTVERDYAGTQFAG--TDYMTTRYAFDA-----ADROKSIHAHRS 1312

Qy 77 LW--QYDLTGNPLCTESIDAGRTVTLLNIEGRPLLTATGVIQTRQYETSLPGRLLSV 134
Db 1313 AWTYTDYLFGRQVSTDPDKGTTTVEYDALDRAVSKTDGRGEVLLFEYD---VLGRKTCM 1369

Qy 135 AEQTPEKTSRITERLIWAGNTEAKDHNLA-----GQCVRHYDPAGVTRLESLSLGT 188

Db 1370 WQSA---KFG-ANKLAASFDTLAGQQDQAVRYEGGETGRAY-TQKVTRYDPLKYVTNN 1424
QY 189 --VLQSSOLL---IDTOEANTGDN--ETVMQ-----NMLADDIVTLTTFDQATGAL 234
Db 1425 ELTLPANDPLVAAGVPARLAPSTGYMLDGTVKQAAAPVAGLSAE---TVSYTYDGLGOV 1481
QY 235 LPOTDAKGNIOBLAYDVAGQNGSMLTLKGQTEQVVIKSLTYSAAGQKLREBHGNDVITE 294
Db 1482 LTAAGTGTGYLQAAASPLGDLRQMTLATDPTGAKKYLNNDEAGTRRLTRSYVTDVHG 1541
QY 295 YSYEPTQRLIGIKTRPSDTKVLQDLRYEYDPCVGNVISIRDAEAETRHQKWPENT 354
Db 1542 F-----MLQELKYQDDAGNITSV-----SDATLLGGTGKA-DHQC 1576
QY 355 YTYDSLYQL-----ISATGEMANIGQOSHQPSPALPSONNTYNTYTRYTYDR 404
Db 1577 FYTDGHRRLSEAWTETADCSISGRVAGLG-----GAAP-----YWTSYQYDD 1620
QY 405 GGNLTKIQHSSPATQNNYTNITVSNRNR- VLSTLTEDPAQVDALFDAGHQNT---L 460
Db 1621 SGLRSK-OTERHMSGDVITTEYEGTAEQPHALSATVTGAENASVYDETCNTETRPGV 1679
QY 461 ISGQMLNWNTRGELQVLT-LVRDKGANDREWRYSGDGRMLKINEQQAASNNATQTV 519
Db 1680 RATOTLDWNAEGLAGVSEPAAGKPGATGATYVYDAGD-----LLIRPTTDTG---ETV 1732
QY 520 TYLPNLELRLTONSTATTEDLQVITVGEAGRAQVRLVHWSCKPEDIDNNQLRYSVDMLI 579
Db 1733 LYLGTTEVHLKVSNGAAKALSARTYKAGSAVIAVRTSTAG-----VSGTKLTFLAGDHH 1788
QY 580 GSSQLELDSGOIISSEEEYYPGGTALWAARNTQEAQYKTIIR-YSGKERD-ATGLYYTYG 637
Db 1789 GTSGLAINADTLAFAKRWSPTFG-----APRGTAGWPDGRFLGKPADATAATGLTOLCA 1843
QY 638 RYQPWIGRWLSSDPAGITD-----GLNLRYMVRNPPVLLDPGLMPTTAERTAAALK--N 692
Db 1844 RQYEDPTGRFLSVDPLEPKDPTNLNGYAYASNSPVTNSDPSGSDGLGLGAGIATIG 1903
QY 693 KYTDSAPSANATVAINIRPPVAPKPSLPKASTSSQPTTHPIGAANIKPTTSGSSIVAP 752
Db 1904 GVGVAGVAGAAITAVG-----SLGGGGGG-----GNGTAPTSSGG----- 1941
QY 753 LSPVGNKSTSEISLPESAQSSSTSTNLQKSFLL-YRADNRSEEMQSKFPEGPKAW 811
Db 1942 -----WTQPLTKQWTFGAT--YNFITKSWDLPPNPPSQSLEMLASMPD---W 1984
QY 812 TPL-DTQMARQF---ASIFIG 828
Db 1985 GIVSDPKAANRWETSRSLFFG 2005

RESULT 10
E90658
RhaG core protein with extension [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90658
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; PMID:21156231; PMID:11258796
A:Accession: E90658
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1404 <HAY>
A:Cross-references: UNIPROT:Q8XED9; UNIPARC:UPI00001653A9; GB:BA000007; PIDN:BA033660.1;
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC80237
C:Superfamily: rhesf protein

Query Match 6.0%; Score 300; DB 2; Length 1404;
Best Local Similarity 21.4%; Pred. No. 9.1e-09;
Matches 192; Conservative 118; Mismatches 341; Indels 246; Gaps 40;
QY 18 YDNRG--LIIRNIDPHRTTANGDPDTRIT-----RHOYDIHGLNQSIDPRL--- 62
Db 646 YDAAGRTVLTNENGSSQSTFRYDPVDRLTEQGFQDRTQRYHYDITGKLTQSEDEGLVTL 705
QY 63 --YEAKQ--TNNTIK--PNFLQWYDLTGNPLCTESIDAGRTVTLN---DIEGRPLLTWTA 113
Db 706 WHYDASDRITHTRVNGDPAEQWQYDEHGWLTTLTSHTSEGRVSVHYGDDKGR-----L 759
QY 114 TVIVOTROYESSLPGRLLSVAEQPEEKTSRITERLWAGNT-EAEKHNLNLA-----G 166
Db 760 TGERQT-----VENPE-----TCMLWEHETGHAYSQGLTAQRPDG 797
QY 167 QCVRHVDTAGVTRLESLSLTGVLSSQLLIDTOEANTGDNVETWQNMMLADDIVTLT 226
Db 798 LPPVEWLYTSGYLAKMKLGGTFLVEYMRDLRHRETARSG-----GEAYELAT 846
QY 227 TPDATGALLTQTDAKGNTQRLAYDVAGQNGSMLTLKGQTEQVVIKSLTYSAAGQKLRE 286
Db 847 AWNTSGQLRSR---HLNLPQLDRDYDNDNGQLIRISGPQES----- 885
QY 287 HGNVITEYSYEPETORLIGIKTRRPSDKVLQDLRYEYDPCVGNVI---SIRDAEATRP 343
Db 886 -----REYRYS-DTGRLTGVHT---TAANLDIDIPATDPAGNRLPDPELHPDSTLTA- 934
QY 344 WHNOKVMPENTVYDSLYQLISATGEMANIGQOSHQPSPALPSONNTYNTYTRYTYD 403
Db 935 WPDNRIADAHVY--RYD-----EYGLAEKTDRIPEGVI-----RMDERTHHYVD 981
QY 404 RGGNL---TKIOHSSPATQNNYTNITVSNRNRVAV-----LSTLTEDPAQVDALF 451
Db 982 SQHRLVFHTRIHQGEPOVESRYLD-PLGRRTGKVRWRERDLTGWMSLSRKPETWTYGM 1040
QY 452 DAGHQNTLISGQNLNWNTRGELQVTLVKRDKGANDREWRYSGDGRMLKINEQQAAS 511
Db 1041 D--GDRLTVQOTQTRIQTVYQPGSFTPLLRITENGEGQAKARH-----RSLAEVLQEDTG 1094
QY 512 NNAQTVRYLPNLELRLTONSTATTEDLQVITVG-----EAGRAQVRLVHW--- 558
Db 1095 VTLPAELAVMLGRLELRQGSVSESOQWLAAQGLTAEQMAQLEAGVYPERKHLHYC 1154
QY 559 -ESGKPEDIDNNQLRYSYDNLIGSSQ-----LELDSGOIISSEEEYYPGGTALWAARNO 612
Db 1155 DQGLPL-----GLISPGRETALAEYDEWGNLLSETSAQPL----- 1191
QY 613 TEASYKTIYSGKERD-ATGLYYGYRYQWIGRWLSSDPAGITDGLNLRYMVRNPPVT 671
Db 1192 ----QOQLRFPQQYDEESGLYTNRRNYDPLQGRYITQDPTGLEGGWNLVYPL-NPTE 1246
QY 672 LLDPDGLMPTTAERIAALKKKNKVTDSAPSANATVAINIRPPVAPKPSLPKASTSSOPT 731
Db 1247 HIDPUGL-----ALDNLVYSDPIYKGLNV-----REFPT 1278
QY 732 THPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSS-----TTS 779
Db 1279 GFTVGHG-SPTSMSSDDRI-----KGSDLTIKQLASDIRANPKYHEGMPVVLFSCE 1329
QY 780 TNLQKSKSTLYRADNRSP-----EMQSKFPFGFVAVTPLDTYMARQPAISIFIGKD 831
Db 1330 TCKGKNSPAQKLANELDATVIAPDEIITWPDGNYA-----IMGQTAIRITIGKD 1379

RESULT 11

A10093
conserved hypothetical protein YPO0762 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: A10093
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic
 A:Reference number: A96629; MUID:21156231; PMID:11258796
 A:Accession: F91187
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1409 <HAY>
 A:Cross-references: UNIPROT:Q8XD12; UNIPARC:UPI000015551E; GB:BA000007; PIDN:BAB37893.14
 A:Experimental source: strain O157:H7, substrain RIMD 0509552
 C:Genetics:
 A:Gene: ECs4470

	Query Match	5.7%	Score 285;	DB 2;	Length 1409;
	Best local Similarity	20.4%;	Pred. No. 6.5e-08;		
	Matches 209;	Conservative 125;	Mismatches 297;	Indels 396;	Gaps 54;
Qy	18	YDRGLIIRNIDFHRITTTANGDPDTRTRHGYDTHGHLNQSIDPRLYEAKQTNNIKPNFL	77		
Db	575	YDSRGQII-----AVKDTQGHETRYYNAGDLTTVIAP-----DGSRNGT-----	615		
Qy	78	WQYDLTGNPLCTESIDAGRTVTLLNDIEGRPLLTATGTATGVIQTRQYETSSLPGRLLSVAEQ	137		
Db	616	-QYDANGAKAICT-----TQGLTRSMEXDAA---GRVIRLTSE	649		
Qy	138	TPSEKTSR--ITERLI-----WAGNTEAKDHNLAGQCVR-----HYDTAGVTIRLE	181		
Db	650	NGSHITPRYDVLDELIOETGFDGRTQ-RYHHDLTGKLIIRSEDEGLVTHWHYDEA---DRLT	706		
Qy	182	SLSLTGTV-----LSQSOLLID---TOEANWTGDNETV-----	212		
Db	707	HRTVKGTAEKQYDERGWLTDISHSEGRVTVHYGYDEKGLTGERQIVHHQPTEALL	766		
Qy	213	WQNMLADDIYTLTSTDATG-ALLTQTDAGNQLAYD---VAGQLNGSWLTKGQTEQ	268		
Db	767	WQ-----HETRAYNAQGLANRCIPDSLPAVEWLTYGSGWLAGMKLGDTPLVDFTDR	819		
Qy	269	VIKSL-----TYSAGQKUREHGNV-----ITEYS	296		
Db	820	LHRKTLRRFRGRVELTAYTPAGQ-LSQHLNLSQYDRDYTWNDNGELIRISSPQTRSYS	878		
Qy	297	YEPEQRLIGIKTRPSDTKVLQDLREY--DPVGVI-----SIRNDAAEAFRWHNOKVMP	351		
Db	879	YS-DSGRLTGVT-----TAANDIRIPYATDPAGNKLDPPELHPDSTLS-MWPDNRIR	931		
Qy	352	ENTVYDLSYQLISATGREMANICQOSHQPSPALPSPDNNTYNTYRTTYIDRGNNL---	408		
Db	932	DAHLY-----RYDRGR-----LTEKDLPIEGVIRTDDE---THRYHYDSQHLRVHY	978		
Qy	409	TKIOHSSPATQNNYTNITVSNRSRAVLSTLTEDPAQVDALFDAG-----HQNLT	460		
Db	979	TRTQYEEPLVESRY-----LYDPLGRRVAKRVWRERD	1011		
Qy	461	ISGNLWNATRGELQVTLVKRDGANDREWYRSGDGRMLKINEQQASNNAAQTORVT	520		
Db	1012	LTG-----WMSLSRKPKQT-----WYGDGD--RLTTI-----QNDKTRIQTI	1047		
Qy	521	YLPNLEURLTQNSTATTEDLQVITVGAGRAQVRVLH---WESGKPED-----	565		
Db	1048	YQGSFTPLLRVETAT-----GELAKTORSLADALQSGGEGDGGSVFPPLVQM	1098		
Qy	566	IDNNQLFRSYDNLIGSSQLELDSGQIISBEE-----YYP-----YGG--TALW	607		

```
Db 1099 LDRLESEILADRVSESRWLASCLGTVAQMSQMDPVYTPARKTHLYCHDRGLPLALI 1158
QY 608 AARNOTE--ASY-----KTIYSKGERD-ATGLYYGYVRYQPWIGRWL 648
Db 1159 SAEGATENCAEYDEGNLLNBEENPHQLQQLRLPGQQYDEESGLYNNRHYDPLHGRI 1218
QY 649 SSDPAGTIDGLNLYRMVRNPNVTLDPDGLMPTIAERIAALKKVKVTSAPSANATVA 708
Db 1219 TQDPILGLKGGWNYQYPL-NPVINDPQOL-----VD 1249
QY 709 INIRP-----PVAKPSLPRKASTSSQPTTHPIGAANKPTTSGSSIIVAPLSPVGNKSTS 762
Db 1250 INLYPESDLIHSVADEINIPGVFT-----IGHGHT- 1279
QY 763 BISLPESAQSSSSSTSTNLQKSTLYRADNRSPFEMQSKPPEGFKAWT-PLDT-KMAR 820
Db 1280 ----PTSIESATRSIWTAK-----DLAYLIKFDGNYKDGWTLVWSCNTGKQON 1324
QY 821 QPASFIFGQKQTSNLPKBTVRKNISTW---GAKPKL-----KD 854
Db 1325 SPASQAKELHTNVIGPDT---LWTWGRGTNGKLMKMDVLTAPTNLNSKOLMAITTKD 1381
QY 855 LSNYIKY 861
Db 1382 LGNWITY 1388

RESULT 13
C64805
rhesC protein precursor [similarity] - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C>Date: 12-Sep-1997 #sequence revision 17-Sep-1997 #text_change 09-Jul-2004
C:Accession: C64805; 169400; T48912
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: C64805
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1397 <BLAT>
A:Cross-references: UNIPARC:UPI00001338AF; GB:AE000173; GB:U00096; NID:6
A:Experimental source: strain K-12, substrain MG1655
R:Peulner, G.; Gray, J.A.; Kirschmann, J.A.; Lehner, A.F.; Sadosky, A.B.; Vlazny, D.A.;
J. Bacteriol. 172, 446-456, 1990
A:Title: Structure of the rhesA locus from Escherichia coli K-12 and comparison of rhesA w
A:Reference number: 154935; MUID:90094253; PMID:2403547
A:Accession: I69400
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1221-1397 <RES>
A:Cross-references: UNIPARC:UPI000017888F; GB:M29718
A>Note: the sequence is revised in GenBank entry ECORHSCA, release 114, (PIDN:AACG3073.1
R:Oshima, T.; Aiba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Honjo, A.; Ikemoto, K.; Ihada
T.; Mizobuchi, K.; Mori, H.; Motomura, K.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Sai
DNA Res. 3, 137-155, 1996
A:Title: A 718-Kb DNA Sequence of Escherichia coli K-12 Genome Corresponding to the 12.7
A:Reference number: 225006; MUID:97061202; PMID:8905232
A:Accession: T48912
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1397 <OSH>
A:Cross-references: UNIPARC:UPI00001338AF; EMBL:D90709; NID:61651305; PIDN:BAJ35359.1; F
A:Experimental source: strain K12; Kohara clone 174
C:Genetics:
A:Gene: rhesC
A:Map position: 15.7-16.0
C:Superfamily: rhesF protein
Query Match 5.7%; Score 283; DB 2; Length 1397;
Best Local Similarity 21.3%; Pred. No. 8.3e-08;
```

```
Matches 175; Conservative 96; Mismatches 242; Indels 308; Gaps 43;
QY 18 YDNRLIIRNIDPHRTTANGDPDTRITRHQYDIHGLNQSIDPRLYEAKQNTNTIKPNFL 77
Db 575 YDSRGOLI-----AVKDTQGHETRYEYNAAGDLTVIAP-----DGSRNGT----- 615
QY 78 WQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTRQVETSSLCGLLSVAEQ 137
Db 616 -QYDANGKAITC-----TQGLTRSMEDYDAA---GRVIRLTSE 649
QY 138 TPEEKTSR--ITERLI---WAGNTEAEKDHNLACQVR-----HYDTAGVTRLE 181
Db 650 NGSHTTFYDVLDRLIQETGFGRTQ-RYHDDLTKLIRSEDEGLVTHWHYDEA---DRLT 706
QY 182 SLSLTGTVLSOSSQLLIDTQBANWGD----- 208
Db 707 HRTVNGETAERWQY-----DERGMLTDISHSEGHRVTVHYGDSKGRLASBHLTVHHQ 761
QY 209 -NETVQWNLADDIYTTLSFTDAGT-ALLTQTDAGNIQRLAYDVAGOLNGSWLTKGOT 266
Db 762 TNELLWQ-----HETRHAYNAOGLANRCIPDSLPAVEWLTYG-SGWLSGMKL---GDT 810
QY 267 EOY-----IYKSL-----TYSAAAGOKLREEHGNVDITE----- 294
Db 811 PLVEYTRDLRHRETLRSFGRYELTAYTPAQ-LQSOHLNLSLSDRDYTDWNGELIRIS 869
QY 295 -----YSYBPETQRLIGIKTRRPSDTKVLQDLRYEY--DPVGNVISIRNDAE-----A 340
Db 870 SPQRTSYSYS--TTCRLTGVT-----TAANLDIRIPYTTDPAGNRLP---DPHLHPDSA 920
QY 341 TRFHNQKMPENTYDLSYQLISATOREMANIQQSHQSPSPALPDNNNTYTYTRY 400
Db 921 LSWMPDNRIARDAHYLY-----RYDRHGR-----LTEKTDLPIEGVIRTDDE--THRY 967
QY 401 TYDRGSLN---TKLOHSPATQNNYNTTIVSNRSRAVSLTLEDPAQVDALFDAGG-- 455
Db 968 HYDSOHLVHVHTTQYAEPLVESRY-----LYDPLGR 1000
QY 456 -----HONTLISQNLNWNTRGELQVTLVKRKGANDREWYRSGDRMLKINEQO 509
Db 1001 VAKRWREERDLTG-----WMSLSRKPOVT-----WYGMWGDG--RLTTI--- 1037
QY 510 ASNAQATQRTVYLPNLEURLTQNTATTEDLVITVGEAGRAQVRVLH---WESGKPED- 565
Db 1038 -QNDRTIQTQYQGSFPLIRVETAT-----GELAKTQRRSLADTLQOOSGEDGG 1087
QY 566 -----IDNNQLRYSYDNLIGSSQLELDSGQLISBEE-----YYP----- 600
Db 1088 SVMPPFVLVQMLDRLESEILADRVSEESRRWLASCGLTVAQMSQMDPVYTPARKHLYH 1147
QY 601 -----YGGTALWAA-----RNQTEASYKTIYSKGERD-ATGLYYGYG 637
Db 1148 CDHRCGLPLALISTEGTTAWAYAEYDEWGNLLNEENPHQLQQLRLPGQQYDEESGLYNNRH 1207
QY 638 RYQPWIGRWLSSDPAGTIDGLNLYRMVRNPNVTLDPDGL 678
Db 1208 RYYPDQGRYITQDPIGLKGGWNYQYPL-NPISNIDPLGL 1247

RESULT 14
AF1489
cell wall-associated protein precursor wapa (B. subtilis) homolog lin0454 [imported] - i
C:Species: Listeria innocua
C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 09-Jul-2004
C:Accession: AF1489
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecker,
.; Dominguez-Bernal, G.; Duchaud, B.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
```

A:Accession: AF1489
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2167 <GLA>
A:Cross-references: UNIPROT:Q92EKS; UNIPARC:UPI00000CC234; GB:AL592022; PIDN:CAC95686.1;
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin0454

```
Query Match          5.6%; Score 282.5; DB 2; Length 2167;
Best Local Similarity 22.6%; Pred. NO. 1.8e-07;
Matches 168; Conservative 80; Mismatches 279; Indels 215; Gaps 34;

Qy 14 TVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGH-----LNQSDP 60
Db 1476 TKNYDASG-----NYVTASYDEGRKTSFTFYDYGKTSDEKGNKKTLYDADN 1527

Qy 61 RLYEAKQNTNNTIKPNFLWQYDLTGPNLCTESIDAGRTVTLNDIEGRPLLTATAGVQTQR 120
Db 1528 ALIDTKLANGT---SVAYKYD-----DNGNTTEKN-----VTASGKTQKN 1564

Qy 121 QYETSSLPGRLLSVAEQPEKTSRITERLIWAGNTEAEKDHNLGAGCVRH-YDTAGVTR 179
Db 1565 IY-----EYVDVNKITAFTDAL-----NRTIKVEYDAAG-NE 1595

Qy 180 LESLSLTGTVLSSQSOLLIDTOEAWNT-----GDNETWQNLADDI---YTTLS 226
Db 1596 TKAIMPNRGRVTESTYDSADRMGDKWMDKLAFLQFPYDPNGNOT---KVTDEINSIVTDK 1651

Qy 227 TFDATGALLTQTDAGKNIORLAYDVAGOLNGSWLTGKGTQOV-----IISLT 275
Db 1652 TYDDANRITKVAERGGDVSYTKPKYDN-----KGTDKVGEVAINHGDIYAKTSYT 1705

Qy 276 YSAAGQKLRERH-----GNDVITEYSYEPETQRLIGIKTRRPSDT 315
Db 1706 YNDLDRNTRVNDGSKNAYFEFDEFGNINVTAGNGTAANYTD-STQKVTNAAISSASGT 1764

Qy 316 KVLQDLRYEYDPGVNVSIRNDAEATRFWNNQKVPENTTYTDSLYQLISATGREMANIG 375
Db 1765 QIL-DENYTYDAASNRTSIDNKQDG-----KTTYEYDAVNOLTKET----- 1804

Qy 376 QQSHQFSPALPSDNNTYNTYTRTYDGRGNLTKIQHSSPATQNNYTNITVNSNRRA 435
Db 1805 -----LPDGT-----VKAYTYDFGNRTQV-----AISGETKTIDASVNDGNQ 1843

Qy 436 VLSTLTEDPAQVDAL-FDAGGHQNTLISGQNLNNTRGELQOQVTLVKRDKGANDREWYR 494
Db 1844 LVSNWGE-----ALTYDANGNR-TSDGKYTYTWDTGRLSSIT-----KKGSEBPFTSYT 1892

Qy 495 YSGDGRMLKINEQOASN---NAQTQRTYVLPNLELRLTQNSTATTEDLQVITVEAGRA 551
Db 1893 YDDNRRLSKTVDTGVTNTHYDGSIDVLY-----ETDGGKVVRQVYVSDD 1939

Qy 552 QVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSEGQIISBEEYYPVGGTALWAARN 611
Db 1940 NVRLAMKQNGK-----TLTYHY-NAHGQVIALTDGKLVAEYDANGVNLKNTAST 1991

Qy 612 QTEASYKTIRYSGKRD-ATGLYYGYRYQYQFWIGRMLSSDP-AGTIDG-----LNLRYMVR 666
Db 1992 E-EAKANPYGAGTYDKIEIQYLMARYEPEQGVFTAYDPYFCDDEDDPQTWNGYNAN 2050

Qy 667 NNPVTLDDPDGLMPTIARIAA 688
Db 2051 NNPVMMFDPDG---NVAWIIAA 2069
```

RESULT 15
AH0439
probable membrane protein YPO3615 [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Jul-2004
C:Accession: AH0439
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0439
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1512 <KUR>
A:Cross-references: UNIPROT:Q8ZB14; UNIPARC:UPI000000CDRA53; GB:AL590842; PIDN:CAC93084.1;
C:Genetics:
A:Gene: YPO3615

```
Query Match          5.6%; Score 279.5; DB 2; Length 1512;
Best Local Similarity 22.8%; Pred. No. 1.5e-07;
Matches 179; Conservative 106; Mismatches 300; Indels 201; Gaps 43;

Qy 9 YQKTPYTSV-YDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP----- 60
Db 677 YPSQOTWSFRYKKGKGEVLKLI-----IAEQVMRYHYDHHGCLSTIIDPKGNSTAV 727

Qy 61 -----RLYE-----AKQNTNNTIKPNFLWQ-----YDLTGPNLCTES 91
Db 728 TLDVGLRGLFSHQNALGELTRYTHSDAHASPAQSVTKMMPDGVGEQAIAYD-SEKRIAAIT 786

Qy 92 IDAGRTVTLNDIEGRPLLT--VTATGVIQTRQYETSSLPGRLLSVAEQPEKTSRITER 149
Db 787 DGAKT-TRVEYGGFDLLTGLIRPDGQRLTFGYDITLRLNQVTVASGDTYRTYTRDR----- 841

Qy 150 LIWAGNTEAEKDHNLGAGCVRH-YDTAG---VTRLESLSLTGTVLSSQSOLLIDTOEANW 205
Db 842 ---AGQVISETD---FTGRTVHYQYDAVGRIRGARYPQORLVRWHVSMQDQVL---AQQTW 893

Qy 206 TGDNETWQNLADDIYTTLS-TFDATGALLTQTDAGKNIORLAYDVAGQ-----LNG-- 257
Db 894 HCD-----ALSSLTGTVSYGYDGAGRLSATNADAVVE-FDYDEAGOLVAERLANGRE 945

Qy 258 ---SMLTLKQ--TEQVIKSLT--YSAAGQKLEEHGNDVITEYSYEPETQRLIGIKTR 310
Db 946 VRHQMWDALNGTPVARQVEGLTFTVGAQGBELTLQLAGHQPLQLQHD-----RLGRETV 1000

Qy 311 RPSDTKVLQDLRYEYDPVG---NVISIRNDAEATRFWNNQKVPEN-----TITY 357
Db 1001 RESAAGFIQ--ACNYTFSGLLAHQAGNSA---LFQOQLIAPESPALHGSVAVNRSQY 1054

Qy 358 DSLIQLISATGREMANIQQSHQFP--SPALPSDNNTYNTYTRTYDGRGNLTKIQHSS 415
Db 1055 DRAYNVV---GMDDGRWGKTYQYDRNDQVVRADFGGFLPLQEQFYSYDVNQNLRE-HRCL 1110

Qy 416 PATQNNYTNITVNSNRNAV-----LSTLTEDPAQVDALFDAGGHQNTLISGQ 464
Db 1111 PRGAQVLAQASQOQOQAGRVVRGDSQVRYDAAGRLVEKRSQKQ-----GYRQL---W 1161

Qy 465 NLNNTTRGELQOQVTLVKRDKGANDREW-YRYSGDGRRMLKINEQOASNNAAQTQRTVYLP 523
Db 1162 RYRNQEQDQLSELL--TPTGA---RWRYGYDAFGRIRKL-----RVVDTP 1202

Qy 524 NL-ELRLTQNSTATT-----EDLOVITVGEAGRAQVRLHW----- 558
Db 1203 PLNEMDAPSTGPATASLAGYAVLWSGDQLIEVEVPYADGTV--AYEQGIHLYAFPGGLTP 1260

Qy 559 ----ESGKPEDIDNNQLRYSYDNLIGSSQLELDSEGQIISBEEYYPVGGTALW-AARNQT 613
Db 1261 MARYAQKQ-----LHYVVADHLGTPRELNLNCGKQVWASRLSTWQAEQLWQAAEE 1312

Qy 614 EASYKTIRYSGKRD-TGLYYGYRYQYQFWIGRMLSSDPAGTIDGLNLYRMVRNPNVTL 672
Db 1313 DRVSCNLRPAQYADAESGLHYNFRYYDGTGQVLCPPDPIGLEGLNPGYGV-HNPVSW 1371

Qy 673 LDDPGL 678
Db 1372 VDPLGL 1377
```

Search completed: February 16, 2006, 21:45:04
Job time : 26.432 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:23:04 ; Search time 128.785 Seconds
(without alignments)
5259.201 Million cell updates/sec

Title: US-10-754-115-47

Perfect score: 5005

Sequence: 1 MKNIDPKLYQKTPVSVYDN.....DAEISFLTLPLKNVPHKR 960

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5005	100.0	960	2	Q8GF97 PHOLU
2	3300.5	65.9	970	2	Q7N7Y4 PHOLU
3	2879	57.5	938	2	Q8GF91 PHOLU
4	2850.5	57.0	949	2	Q8GFA5 PHOLU
5	2833.5	56.6	938	2	Q7N7Y7 PHOLL
6	2821.5	56.4	959	2	Q7N7X5 PHOLL
7	2768.5	55.3	965	2	Q7MZU2 PHOLL
8	2740	54.8	1043	2	Q7MZV7 PHOLL
9	2740	54.7	1043	2	Q85157 PHOLL
10	2539	50.7	915	2	Q7N7Z1 PHOLL
11	2535.5	50.7	915	2	Q33EP1 PHOLL
12	2386	47.7	936	2	Q7MZ20 PHOLL
13	2216	44.3	1016	2	Q6Q1Y7 XENNE
14	2184	43.6	973	2	Q9F9Z0 XENNR
15	2173	43.4	994	2	Q66A95 YERPS
16	2169	43.3	874	2	Q8ZE71 YERPE
17	2168.5	43.3	978	2	Q6AP55 YERFR
18	2148	42.9	982	2	Q66A38 YERPS
19	2138	42.7	984	2	Q8ZE10 YERPE
20	1990.5	39.8	1014	2	Q93RN9 XENNE
21	1833.5	36.6	952	2	Q8ZAV9 YERPE
22	1831.5	36.6	943	2	Q8D1P5 YERPE
23	1816	36.3	1011	2	Q8ZAV8 YERPE
24	1763	35.2	1045	2	Q665G1 YERPS
25	1763	35.2	1045	2	Q693A0 YEREN
26	1301	26.0	886	2	Q87X45 PSESM
27	1274	25.5	927	2	Q4ZXL5 PSESY
28	1268.5	25.3	940	2	Q87X48 PSESM
29	1267	25.3	956	2	Q4ZYW6 PSESY
30	1266.5	25.3	920	2	Q4ZP60 PSESY
31	1238.5	24.7	923	2	Q4ZTS1 PSESY

32 1235.5 24.7 881 2 Q4ZP54 PSESY Q4ZP54 pseudomonas
33 1223.5 24.4 955 2 Q4ZP55 PSESY Q4ZP55 pseudomonas
34 1217.5 24.3 939 2 Q4ZP59 PSESY Q4ZP59 pseudomonas
35 1215.5 24.3 922 2 Q887R0 PSESM Q887R0 pseudomonas
36 618 12.3 163 2 Q6WCC0 XENNR Q6WCC0 serratia pr
37 590.5 11.8 2439 2 Q4HWU2 GIBZE Q4HWU2 gibberella
38 518 10.3 1140 2 Q73L45 TREDE Q73L45 treponema d
39 517.5 10.3 1488 2 Q73L43 TREDE Q73L43 treponema d
40 492 9.8 1126 2 Q73MB8 TREDE Q73MB8 treponema d
41 479.5 9.6 3320 2 Q73MFO TREDE Q73MFO treponema d
42 470 9.4 103 2 Q8KSP1 XENNE Q8KSP1 xenorhabdus
43 395.5 7.9 1746 2 Q5GZ16 XANOR Q5GZ16 xanthomonas
44 357.5 7.1 1513 2 Q7NY44 CHRVO Q7NY44 chromobacce
45 354.5 7.1 1551 2 Q4LW78 BURK Q4LW78 burkholderi

ALIGNMENTS

RESULT 1

Q8GF97 PHOLU PRELIMINARY; PRT; 960 AA.
AC Q8GF97
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TremBLrel. 26, Last annotation update)
DE TccC3.
GN Name=TccC3;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA fFrench-Constant R.H.;
RA "The tc genes of Photorhabdus: a growing family."
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., fFrench-Constant R.H.;
RA "Genomic islands in Photorhabdus."
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AAC17204.1; -; Genomic_DNA.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RHS repeat; 4.
DR TIGRFAMs; TIGR01643; YD repeat 2x; 1.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
SQ SEQUENCE 960 AA; 107061 MW; 595DE383870C8E9E CRC64;

Query Match 100.0%; Score 5005; DB 2; Length 960;
Best Local Similarity 100.0%; Pred. No. 3.1e-259;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHHQYDIHGLNQSIDP 60
|||||
Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHHQYDIHGLNQSIDP 60
|||||
Qy 61 RLYEAKQTNNTIKNFELWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVQTR 120
|||||
Db 61 RLYEAKQTNNTIKNFELWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVQTR 120
|||||
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGAGQCVRHYDTAGVTRL 180
|||||
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDHNLGAGQCVRHYDTAGVTRL 180
|||||
Qy 181 ESSLTGTVLSCSSQLLIDTQEAQNWGTGNETVQWMLADDIYTTLTSTFDATGALLTQIDA 240
|||||

```
Db 181 ESSLTGTVLSQSSQLLIDTQEAANWTDGNETVQWNLADDIYTTLTSTFDATGALLTQIDA 240
Qy 241 KGNLQRLAYDVAGQINGSWLTGKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Db 241 KGNLQRLAYDVAGQINGSWLTGKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Qy 301 TORLIGIKTRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFVHWNKWPENYTYDLSL 360
Db 301 TORLIGIKTRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFVHWNKWPENYTYDLSL 360
Qy 361 YQLISATGREMANIQOSSHQPPSPALPDDNNNTYNTYTYDREGNLTQKHSSPATON 420
Db 361 YQLISATGREMANIQOSSHQPPSPALPDDNNNTYNTYTYDREGNLTQKHSSPATON 420
Qy 421 NYTNTITVSNRRAVLSLTLEDPAQVDALFDAGHQNTLISGQNLNWNTRGELQOQVTLV 480
Db 421 NYTNTITVSNRRAVLSLTLEDPAQVDALFDAGHQNTLISGQNLNWNTRGELQOQVTLV 480
Qy 481 KRDKGANDREWYRSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTATTEDL 540
Db 481 KRDKGANDREWYRSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTATTEDL 540
Qy 541 QVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIESEYYP 600
Db 541 QVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIESEYYP 600
Qy 601 YGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYPWIGRWLSSDPAGTIDGLN 660
Db 601 YGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYPWIGRWLSSDPAGTIDGLN 660
Qy 661 LYRVNRNPVTLDPDGLMPTIAERIAALKKVKYTDGAPSPANATVAINRPPVAPKPS 720
Db 661 LYRVNRNPVTLDPDGLMPTIAERIAALKKVKYTDGAPSPANATVAINRPPVAPKPS 720
Qy 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSTTST 780
Db 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPESAQSSSSTTST 780
Qy 781 NLQKSFYLYADNRSFEEMQSKPEPKAWPTLDTKWARQFASIFIGQKDTSNLPKETV 840
Db 781 NLQKSFYLYADNRSFEEMQSKPEPKAWPTLDTKWARQFASIFIGQKDTSNLPKETV 840
Qy 841 KNIISTWGAQPKLKDLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
Db 841 KNIISTWGAQPKLKDLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
Qy 901 GQKLNPLPEGRTKNMPVSLLLDTFQIETSSIIALNHGVPNDABISFLTTPKKNVQPKHR 960
Db 901 GQKLNPLPEGRTKNMPVSLLLDTFQIETSSIIALNHGVPNDABISFLTTPKKNVQPKHR 960
```

RESULT 2

```
Q7N7Y4 PHOLL
ID Q7N7Y4 PHOLL PRELIMINARY; PRT; 970 AA.
AC Q7N7Y4;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein Tccc3.
GN Name=tccc3; OrderedLocusNames=plu0967;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1];
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
EX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt.886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-F.,
RA Dassa E., Deroose R., Derzelle S., Freyssiuet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
```

```
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
luminescens.";
RL Nat. Biotechnol. 21:1307-1313(2003).
DR EMBL; BX571862; CA613262.1; -; Genomic_DNA.
DR Photolust; plu0967; -.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; RHS repeat; 5.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 1.
KW Complete proteome.
SQ SEQUENCE 970 AA; 109550 MW; C62D47037EAD644A CRC64;

Query Match 65.9%; Score 3300.5; DB 2; Length 970;
Best Local Similarity 67.6%; Pred. No. 5.5e-168;
Matches 679; Conservative 81; Mismatches 151; Indels 93; Gaps 17;

Qy 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRHTTANGDDPDRITRPHQYDIGHILNOSIDP 60
Db 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRHTTANGDDPDRITRPHQYDIGHILNOSIDP 60
Qy 61 RLYEAKOTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
Db 61 RLYEAKOTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLAGOCVRRHYDTAGVTRL 180
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLAGOCVRRHYDTAGVTRL 180
Qy 181 ESSLTGTVLSQSSQLLIDTQEAANWTDGNETVQWNLADDIYTTLTSTFDATGALLTQIDA 240
Db 181 ESSLTGTVLSQSSQLLIDTQEAANWTDGNETVQWNLADDIYTTLTSTFDATGALLTQIDA 240
Qy 241 KGNLQRLAYDVAGQINGSWLTGKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Db 241 KGNLQRLAYDVAGQINGSWLTGKQTEQVVIKSLTYSAGOKLREEHGNDVITEYSYEPE 300
Qy 301 TORLIGIKTRR - - - - - PSDTKVQLDLYEYDVPVGNVISIRNDAEATRFVHWNKWPENY 355
Db 301 TORLIGIKTRRSLDTPNPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFVHWNKWPENY 360
Qy 356 TYDSLQYLIATGREMANIQOSSHQPPSPALPDDNNNTYNTYTYDREGNLTQKHSS 415
Db 361 IYDSLQYLIATGREMANIQOSSHQPPSPALPDDNNNTYNTYTYDREGNLTQKHSS 420
Qy 416 PATQNNYNTNITVSNRRAVLSLTLEDPAQVDALFDAGHQNTLISGQNLNWNTRGELQ 475
Db 421 PATQNNYNTNITVSHQSNRAVLSLTLEDPAQVDALFDAGHQNTLISGQNLNWNTRGELQ 480
Qy 476 QVTLVKRDKGANDREWYRSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTA 535
Db 481 QVTLVKRDKGANDREWYRSGDGRMLKINEQOASNNATQORVYLPNLELRLTONSTA 540
Qy 536 TTEDLQVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIESE 595
Db 541 TTEDLQVITVGEAGRAQVRLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIESE 600
Qy 596 BEEYYPGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYPWIGRWLSSDPAGT 655
Db 601 BEEYYPGGTALWAARNQTEASYKTIYSGKERDATGLYYGYRYQYPWIGRWLSSDPAGT 660
Qy 656 IDGLNLYRMVRNPNVTLDPDGLMPTIAE - - - - - RIAALKKNKVTDS 697
Db 661 IDGLNLYRMVRNPNVTLDPDGLMPTIAE - - - - - RIAALKKNKVTDS 718
Qy 698 APSANATVAINRPPVAPKPSLPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPV 757
Db 719 - - - - - IDEYNAGLIMGISLSYDFRISQNLKEDSPPKLNENIR - TNSSKYLVA - - - - - MG 769
Qy 758 - NKSTSEISLPESAQSSSSTTSTNLQKSFYLYADNRSFEEMQSKPE - - - - - 805
Db 770 IEESATDKIKQEAQEYQAYTA - - - - - SWSFHWENKGLDIQSKVKNLPSKLSFG 823
Qy 806 -EGKAWTPTDKWARQFASIFIGQKDTSNLPKETVKNKIST-WGAKPKLKDLSNYIKYTK 863
Db 806 -EGKAWTPTDKWARQFASIFIGQKDTSNLPKETVKNKIST-WGAKPKLKDLSNYIKYTK 863
```

```
Db 824 LSGQRFOKLD-----AFILKNSDGSDFTIQORDBILTKFASAGNTELESTLS-AA 874
Qy 864 DKSTVWSTAINTEAGSGSSAPLHKIDMDLY-----EF-----AIDGOKL----- 904
Db 875 DQS--WLKNTATATAPRQTS-----KLGDWFIKQLASPDFRFVWAGYNGEKLITIEQLKS 927
Qy 905 -NLPPEGTKNNVPSLLDTPQIETSSIIALNHGPNVNDABISFL 947
Db 928 EOPWTKGRRKEGPN---EYAEAITFSEIRHAYRKKYDSTINFI 968

RESULT 3
Q8GF91 PHOLU PRELIMINARY; PRT; 938 AA.
AC Q8GF91
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Tccc5.
GN Name=Tccc5;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_taxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
RT "Genomic islands in Photorhabdus.";
RL Trends Microbiol. 10:541-545(2002).
DR EMBL; AF346500; AA017210.1; -; Genomic_DNA.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR006530; YD.
DR Pfam; PF05593; Rhs_repeat; 4.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN 1.
DR SEQUENCE 938 AA; 105094 MW; 080127894806940B CRC64;

Query Match 57.5%; Score 2879; DB 2; Length 938;
Best Local Similarity 60.9%; Pred. No. 1.9e-145;
Matches 591; Conservative 90; Mismatches 186; Indels 104; Gaps 16;

Qy 1 MKNIDPKLYQKPTVSVVNDRLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60
Db 1 MENIDPKLYHPTVSVVNDRLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60
Qy 61 RLYEAKQTNNTIKPNFLWOYDITGNPLCTESIDAGRTVTLNDIEGRPLLTIVTATGVIQTR 120
Db 61 RLYDAKQTNNAVQPNFIWHNLITGNILATESVDAGRTITLNDIEGRPLLTIVTATGVIQTR 120
Qy 121 QYETSSLPGRLLSVAEQ-TPEKTSRITERLWAGNTEAEKHNLAGQCVRHYDAGVTR 179
Db 121 RYEDNTLPGRLLAISEQQAEBK--TERLWAGNTPQKHNLAGQCVRHYDAGVTR 177
Qy 180 LESLITGTVLSQSOLLIDTQEAANTGNETWQMLADDIYTLTFDATTGALLTQTD 239
Db 178 LNSLALTGAVLSQSOLLIDTQDADTGDQSLWQKLSSDYVITQSNATDATTGALLTQTD 237
Qy 240 AKNGTQRLAYDVAGOLGNSWLTGKQTEQVIKSLTYSAGOKLREHGNNDVITEYSYEP 299
Db 238 AKNGTQRLAYDVAGOLGNSWLTGKQTEQVIKSLTYSAGOKLREHGNNDVITEYSYEP 297
Qy 300 ETQRLIGIKTRRPSDKVLQDLRYDYPVGNVISIRNDAEATRFWRNOKVAPENSTYDTS 359
```

```
Db 298 ETQRLIGIKTRRPSDKVLQDLRYDYPVGNVISIRNDAEATRFWRNOKVAPENSTYDTS 357
Qy 360 LYQLISATGREMANIQSHQPPSPALPSDNNNTYNTYTRTYDREGNLTKEHSSPAQ 419
Db 358 LYQLISATGREMANIQSHQPPSPALPSDNNNTYNTYTRTYDREGNLTKEHSSPAQ 417
Qy 420 NNYTNIITVSNRSNRVAVLSTLTEDPAQVDALPDAGCHQNTLISGNLWNTFGELOQVTL 479
Db 418 NNYTNIITVSNRSNRVAVLSTLTEDPAQVDALPDAGCHQNTLISGNLWNTFGELOQVTL 474
Qy 480 VKRDKGANDD--REWYRSGDGRRLKINEQOASNAQTORVYLPNLELRLTQNSTATT 537
Db 475 -----ANNSAGNEWIRYDSNGIRQLKVNEQQTQNIPOQORVYLPGLIEIRTNQNTATT 528
Qy 538 EDLQVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISEE 597
Db 529 EELHVITLKGAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISEE 588
Qy 598 YYPYGGTALWAARNQTEASYKTIYRSGKRDATGLYYGYRYQPWIGWLSDDAGTID 657
Db 589 YYPYGGTALWAARNQTEASYKTIYRSGKRDVTLGLYYGYRYQPWIGWLSDDAGTID 648
Qy 658 GLNLYRMVRNNPVTLDDPDGLMPTTAERIAALKKQKVTDSPANATNVAI-----N 710
Db 649 GLNLYRMVRNNPVTLDDPDGLMPTTAERIAALKKQKVTDSPANATNVAI-----N 710
Qy 711 IRPPVAPKPSLPK-----ASTSQP-----TTHPIGAANIKP-----TTSQSSIVAPL 753
Db 688 MEEAVYKQWAKPQTFKQRAIAAQAQTEQAEHSLTNP--SVDISPIKNTYTTDSSQINAAI 745
Qy 754 SPVGNKSTSEISLPESAQSSSTSTNLQKKSFTLYRADNRSEFEMOSKPFEGKAWTP 813
Db 746 RE--NRITPAV---ESLDATLSSLDQRQWRTYRVMTYVDNS-----TP 784
Qy 814 LDTKWARQFASIFIGQKDTSNLPKTKVKNISWGAQPKLKDLSNYIKYTKDKSTVWVSTA 873
Db 785 SPWHSQPGNSINVGDIIVSDNAYLSTSAH-----RGFLNVHKKETSETRYVQWA 834
Qy 874 INTAGQSGGAPLHKIDMDLYEFAID-----GQKLNPLPEGRTKMVPVSLLDTPQ 925
Db 835 FLTNAGVNVVPAASMYNNAGEQVFKMDLNDRSKSLAEKLLKRVSGPQSQABILLPRETQ 894
Qy 926 IETSSIIALNH 936
Db 895 FE---VWSMKH 902

RESULT 4
Q8GFAS PHOLU PRELIMINARY; PRT; 949 AA.
AC Q8GFAS
DT 01-MAR-2003 (TREMELrel. 23, Created)
DT 01-MAR-2003 (TREMELrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Tccc4.
GN Name=Tccc4;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_taxID=29488;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RA Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RT "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=W14;
RX MEDLINE=22454990; PubMed=12564983; DOI=10.1016/S0966-842X(02)02463-0;
RA Waterfield N.R., Daborn P.J., french-Constant R.H.;
```



```
Db 418 NNYTTNITVSNYSNRAVLSTLTEDPTQVDALFSGGHQTNLSGQVLWTTPRGELKQV-- 475
Qy 480 VKRDGDANDREWYSGDGRMLKINEQOANNAQTORVTVPLNLELRLTQNSTATTED 539
Db 476 -----NSSAGNWHYDNGTFLQKVNQEQTONIAQQQVTVPLGLELRTTQHGTTTIEY 530
Qy 540 LQVITVGBAGRAQVRLHWESKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEEY 599
Db 531 LQVITLKGAGRAQVRLHWESKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEEY 590
Qy 600 PYGGTALWAARNQTEASKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 659
Db 591 PYGGTALWAARNQTEASKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 650
Qy 660 NLYRMVRNPNVTLDPDGLMPTIAERIALKKNKVTDSAPSANATNVAI-----NIR 712
Db 651 NLYRMVRNPNVTFDQGL-----SPANTEEAIKQGSFTGME 689
Qy 713 PPVAPKPSLPKA-----STSSQPTTHPIGAAN-----IKPTTSGSSIVAPLSPVGN 758
Db 690 EAVYKMAKPTQFKQRAIATQEQEAHQLLTNNPGVDTSPDKYTTDSSQI-----N 742
Qy 759 KSTSE---ISLPESAQSSSTSTNLQKSTFLRADNRSEENQSKPEPFKAWTPLD 815
Db 743 TAIRENRITISVKKOLDSSLSALQDRQIRVYRVMTYIDN-----SK-----PSP 786
Qy 816 TKMARQFASIFIGQKDTNLPKPKETVKNISTWGAQKPLKDLNSYIKYTKDKSTVMVSTAIN 875
Db 787 WHSPQEGNSINVDIVSDNAYLSTSAH-----RGFLNFWHKKSTSETRYVKMAFL 836
Qy 876 TEAG-----GSSGAPLHKIDM-DLYEPAIDGQKL--NPLPEGRTKMVPSSLDDT 923
Db 837 TWTGVNVAASKYNNENAEKIFKMLDDSWKSVFVKLIRANGPBAQAEILFPR---ET 893
Qy 924 POIETSSIALNHGPNVDAEISFLTIPLKNVCP 957
Db 894 P-----FVYSMRHQRD-----TYVLLQDIKP 916

RESULT 6
Q7N7X5 PHOLL PRELIMINARY; PRT; 959 AA.
ID Q7N7X5 PHOLL PRELIMINARY; PRT; 959 AA.
AC Q7N7X5;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TccC4.
GN Name=tccC4; OrderedLocuNames=plu0976;
OS Photorhabdus luminescens (subsp. laumondii).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=TT01;
RX MEDLINE=22957627; PubMed=14528314; DOI=10.1038/nbt886;
RA Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A.,
RA Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.-P.,
RA Dassa E., Derose R., Derzelle S., Freysinet G., Gaudriault S.,
RA Medigue C., Lanois A., Powell K., Siguer P., Vincent R., Wingate V.,
RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;
RT "The genome sequence of the entomopathogenic bacterium Photorhabdus
RT luminescens."
RL Nat. Biotechnol. 21:1307-1313 (2003).
DR EMBL; EX571862; CAB1327.1; -; Genomic_DNA.
DR PhotoList; plu0976; -.
DR InterPro; IPR002345; Lipocalin.
DR Pfam; PF05593; Rhs repeat; 2.
DR TIGRFAMs; TIGR01643; YD repeat_2x; 2.
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN_1.
KW Complete proteome.
```

```
SQ SEQUENCE 959 AA; 108232 MW; 1482133DBB080EB7 CRC64;
Query Match 56.4%; Score 2821.5; DB 2; Length 959;
Best Local Similarity 67.0%; Pred. No. 2.4e-142;
Matches 550; Conservative 74; Mismatches 132; Indels 65; Gaps 7;
Qy 1 MKNIDPKLYOKTPTVSYVDNREGLIIRNIDFRTTANGDPDTRITRHOVDIHLNQSIDP 60
Db 1 MKNIDPKLYOHTPTISVYDNRGLTIRNIDFIRSVAGSDTDRITRHOVDHLSQSIDP 60
Qy 61 RLYEAKQNTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLTIVTATGVIQTR 120
Db 61 RLYDAKQNDNSINPNFLWQHLNLTGTLTSEADAGRTVALNDIEGRVLTIVTATNAIQTR 120
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLNLAGQCVRYHYDTAGVTRL 180
Db 121 QYEANTLPGRLLSVSEQAAGEKIPLVTEBHFHWAGNTQTEKHNLNLAGQVRYHYDTAGVTRL 180
Qy 181 ESLSLTGTVLSSQQLLIDTQANWGTGNETVQNMALADDIYTLTSTEDATGALLTOTDA 240
Db 181 KSLSLTENILSSQRLQADGQADWGTNNDPLMQKLSSEIYTTQSTFDATGALLTOTDA 240
Qy 241 KGNIQRLAYDVAGOLNGSWLTLKGTEQVVIKSLTYSAGQKLRREHNDIVITEYSPE 300
Db 241 KSNIQRLAYNVAGQLKGSWLTILKQSEQVVIKSLTYSAGQKLRREHNGALTEYSEPE 300
Qy 301 TORLIGIKTRPSPDKVLQDLRYEYDPVGNVISIRNDAEATRFMHNKMPENTVYDLSL 360
Db 301 TLRIGITTRQSDKNKVLQDLRYEHDVPVGNIIIRNDAEATRFMHNKMPENVYDLSL 360
Qy 361 YQLISATGREMANTGQSHQFPSP-ALPSDNTNTYNTTYTDYDRGGLTKIQLHSSPATQ 419
Db 361 YQLISATGREMANTGQSHQFPSP-ALPSDNTNTYNTTYTDYDRGGLTKIQLHSSPASQ 420
Qy 420 NNYTTNITVSNYSNRAVLSTLTEDPTQVDALFSGGHQTNLSGQVLWTTPRGELKQV 479
Db 421 NNYTTDITVSNRSHAVLNSLTPTQVDTLFDAGGHQKLLPGQDLSSNARGELQVTP 480
Qy 480 VKRDGANDREWYSGDGRMLKINEQOANNAQTORVTVPLNLELRLTQNSTATTED 539
Db 481 VNREN--TSDREWYRYGNDGMRQLKVNQEQTGNSTQQQRTVTVPLGLELRTTQNGTTSD 538
Qy 540 LQVITVGBAGRAQVRLHWESKPEDIDNNQLRYSYDNLIGSSQLELSEGOIIEEY 599
Db 539 LHVITVGAAGHAQVRLHWATTPPAGISNNQLRYSYDNLIGSSQLELNEGOIISQEEY 598
Qy 600 PYGGTALWAARNQTEASKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 659
Db 599 PYGGTALWAARNQTEASKTIRYSKGERDATGLYYGYRYQVWPWGRWLSADPAGTIDGL 658
Qy 660 NLYRMVRNPNVTLDPDGLMPTIAERIALKKNKVTDSAPSANATNVAINIRPPVAPKP 719
Db 659 NLYRMVRNPNITLYDNDGLAP-----PRVRPRH 686
Qy 720 SLPKASTSSQ-----PTTH---PIGANIKPTT-----SGSSIVAPLSPVG 757
Db 687 PDSEGESEDESSIGYRALRSDHEPSIHGLPPPEGANLNISAYAHVRACTSAKVSKSWIS 746
Qy 758 NKSTSEISLPESAQSSSTSTTNLQKSTFLYRADNRSE 798
Db 747 YSRSLKVAASWASGSGSRVVKFRKK-----NNRSFD 779

RESULT 7
Q7MZU2 PHOLL PRELIMINARY; PRT; 965 AA.
ID Q7MZU2 PHOLL PRELIMINARY; PRT; 965 AA.
AC Q7MZU2;
DT 01-MAR-2004 (TReMBLrel. 26, Created)
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TccC6.
GN Name=tccC6; OrderedLocuNames=plu4182;
OS Photorhabdus luminescens (subsp. laumondii).
```


Qy	241	KGNTORLAYDVAGQLNGSWLTLKGTEQVILIKSLTYSAAGQKLREHGHNDVITREYSYPE	300
Db	241	KGNTORLAYDIAGQLKGSWLTVMKQREQVIVKLSWSAAGHKLREHGHNGVITEYSYPE	300
Qy	301	TORLIGIKTR-----PSDTKVLQDLRYEYDVPGNVISIRNDAEATRFWHNOKMPENTYT	356
Db	301	TORLWGLITTRAEBSQSEARVLQDLRYEYDVPGNVISIHNDAEATRFWRNQKVEPENRYV	360
Qy	357	YDSLQILISATGREMANIGQOSHQPSPAL--PSDNNTVTNTVTYDVRGGLNFKIOHS	414
Db	361	YDSLQILMSATGREMANIGQSNQLPSPAMPVPADNSTVTNYQTYSYDVRGGLNVOIRHS	420
Qy	415	SPATONNYTNTIVSNRNRVSLTLTBDPAQVDALFDAGGHQNTLISQNLNWNTRGEL	474
Db	421	SPATONSYTEMTVSSRNRVTLSTLTDTQVDGLDFDAGGHQKTLIPQNLNDNNRGEL	480
Qy	475	QOVLTVKRDKANDREWRYSGDGRMLKINEQOASNNACQORTVYLPNLRLTQNST	534
Db	481	QRVTPVSKEN--SSDSEWRYSSDGRVLRLKVSQQTGSTQVORYTYLPGLRLITGVAD	538
Qy	535	ATTEDLOVITVEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLKSGSOLELDSGQOITS	594
Db	539	KTTQNLQVITVEAGRAQVRVLHWSGKPADIDNNQVRYSYDNLKSGSOIELDSGQILIS	598
Qy	595	EEYYPPGGTALWAARNQTEASYKTIIRYSGKERDATGLYYGYRYQYPWIGLWSSDPAG	654
Db	599	QEEYYPPGGTAVAAKNQTEASYKFIIRYSGKERDATGLYYGYRYQYPWVGWLMSADPAG	658
Qy	655	TIDGLNLYRMVNRNPNVTLDPDGLMPT	681
Db	659	TVDGLNLYRMVNRNPNITLTDHIDGLAPS	685

RESULT 9
O85157 PHOLU PRELIMINARY; PRT; 1043 AA.
ID O85157 PHOLU PRELIMINARY; PRT; 1043 AA.
O85157;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-FEB-2005 (TREMBLrel. 29, Last annotation update)
DE Insecticidal toxin complex protein Tccc (Toxin complex protein).
GN Name:tcctc;
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=23498;
OX [1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=W-14;
RC MEDLINE=98306208; PubMed=9641921; DOI=10.1126/science.280.5372.2129;
RX Bowen D., Rocheleau T.A., Blackburn M., Andreev O., Golubeva E.,
RA Bharti R., ffrench-Constant R.H.;
RT "Insecticidal toxins from the bacterium Photorhabdus luminescens.";
RL Science 280:2129-2132(1998).
RL [2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=W-14;
RC Bowen D.J., Rocheleau T.A., Blackburn M., Andreev O.,
RA ffrench-Constant R.H.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RL [3]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=W14;
RC MEDLINE=21185117; PubMed=11286884; DOI=10.1016/S0966-842X(01)01978-3;
RX Waterfield N.R., Bowen D.J., Fetherston J.D., Perry R.D.,
RA ffrench-Constant R.H.;
RL "The tc genes of Photorhabdus: a growing family.";
RL Trends Microbiol. 9:185-191(2001).
DR EMBL; AF047028; AAC38630.1; -; Genomic DNA.
DR EMBL; AF346499; AAL18473.1; -; Genomic DNA.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0046677; P:response to antibiotic; IEA.
DR GO; GO:0046677; P:response to antibiotic; IEA.

DR	InterPro; IPR002345; Lipocalin.	Query Match	54.7%; Score 2740; DB 2; Length 1043;
DR	InterPro; IPR006530; YD.	Best Local Similarity	74.7%; Pred. No. 6.2e-138;
DR	TIGRFAMs; TIGR01643; YD repeat 2x; 1.	Matches 513; Conservative	78; Mismatches 88; Indels 8; Gaps 3
DR	PROSITE; PS02113; Lipocalin; UNKNOWN 1.		
SQ	SEQUENCE 1043 AA; 111694 MW; E467533FF34C548D CRC64;		
Qy	1	MKNIDPKLYQKTPTVSVYDNRGLIIRNIDFRRITTANGDPDTRITRHHQYDIHGLHNSIDP	60
Db	1	MSPSETTLYTQTPTVSVLDNRGLSIRDIGPHRIVIGGDTDRVTRRHHQYDARGHLNYSIDP	60
Qy	61	RLYEAKQTNTIKNFLWQYDLTGNPCLTESIDAGRTVTLNDIEGRPLLTVTATGVITR	120
Db	61	RLYDAKQADNSKFNFWQHDLAGHAULTESVDAGRTVALNDIEGRSVMTMATGVRQRT	120
Qy	121	QYETSSLPGRLLSVAEOTPEEKTGRITERLLTWAGTEAEKHNLAGQCVRHYDITAGVT	180
Db	121	RYEGNTLPGRLLSVSEQVFNQESAKVIERFIWAGNTTSEKEYNLSELGCIIRHYDITAGVT	180
Qy	181	ESLSLTGTVLSQSSQLLIDTQEAANNWTDGNETVWQNMADDIYTTLTSTEDATGALLTQ	240
Db	181	MSQSLAGAMLQSQHLLAEGQEAANNWSDDETVMQGLASEVYTTOSTTNAIGALLTQ	240
Qy	241	KGNITRLAYDVAGOLNSWLTGKQTEQVILKSLTYSAAQOKLREHGNVDLITEYSYE	300
Db	241	KGNITRLAYDIAGOLKGSWLTVMQOSEQVIVKLSWSAAGHKLREHGNVGVITEYSYE	300
Qy	301	TQRLIGIKTRR----PSDTKVLQDLRYEYDVPGNVISIRNDAEATRFWHNQKMPENTYT	356
Db	301	TQRLIGITTRAEQSQSGARVLQDLRYKYPVGNVISIHDAEATRFWRNQKVEPENRYV	360
Qy	357	YDSLIIQLISATGREMANIQQSHQFPSPAL--PSDNTYNTNTRYTYTDRGGLTKIQHS	414
Db	361	YDSLIIQLMSATGREMANIQQSNNQLPSPIVPPTDDSTYTNLYRTYTYDRGGLVQIRHS	420
Qy	415	SPATONNTYNIIVSNRSRAVLSTLTBDPAQVDALFDAGGHONTLISQNLNWNTRGEL	474
Db	421	SPATONSYYTDDITVSSRSRAVLSTLTDPTRVDALFDGGHOKMLIPQNLDMNIRGEL	480
Qy	475	QQVTLVKRDKGANDREWRYSGDGRMLKINQOASNNAAQTRVTYILPNLELRTQNST	534
Db	481	QRVTPVSREN--SSDSEWRYSSDGMRLLVKVSQQTGKSTQVRVTYLPGLERLTGVAD	538
Qy	535	ATTEDLVITVGEAGRAQVRVLHWSGKPEIDINNOLRYSDYNLTGSSOLELDSQGIIS	594
Db	539	KTTEDLVQITVGEAGRAQVRVLHWSGKPTDIDNNQRYSDYNLTGSSOLELDSQGIIS	598
Qy	595	EEEYYPGGTALWAARQTEASYKTRYSGKERDATGLYYGYRYQYQPWIGRLSSDPAG	654
Db	599	QEEYYPGGTALWAARQTEASYKFIKRYSGKERDATGLYYGYRYQYQPWIGRLSSDPAG	658
Qy	655	TIDGLNLYRMVRNNPVTLLDDPGLMPT 681	
Db	659	TVDGLNLYRMVRNNPITLTDHGLAPS 685	

```

RESULT 10
Q7N7Z1_PHOLL
ID Q7N7Z1_PHOLL PRELIMINARY; PRT; 915 AA.
AC Q7N7Z1;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Insecticidal toxin complex protein TccC2.
DE
DE Name: tccC2; OrderedLocusNames=plu0960;
GN Photorhabdus luminescens (subsp. laumondii).
OS Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Photorhabdus.
OX NCBI_TaxID=141679;
RN [1]

```



```

Db 353 TYIVDSLYQLVSATGREMANAGQGNLPSATAPLPTDSSAYTNTRYRYDRGNLTOM 412
Qy 412 QHSPATQNNYTNITVSNRNRVAVLSTLTEDPAQVDALFDAGGHONTLISQNLNMNTR 471
Db 413 RHPATNNYTTDITVSDRNRVAVLSTLAEPVSDVLMFNGGHOKHLQFQALVWTPR 472
Qy 472 GELQVTLVGRKGDANDREWYRSGDGRMLKINEQOASNAQRTVYLPNLELRLTQ 531
Db 473 GELQVTPVVRDGA-DSESRYDAGSQRIIKTGTQRTGNNVQRTVVVLPGLRLMA 531
Qy 532 NSTATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLGSSOLELDSGQ 591
Db 532 NGVTEKESLOVITVGEAGRAQVRVLHWEIGKPDIDEDSVRYSYDNLVSSOLELDRGY 591
Qy 592 IISSEYYPYGGTALWAARNQTEASYKTIRYSGKRDATGLYYGYRYYPQWIGRLSSD 651
Db 592 LISEEPYPYGGTAVLTARSEVADYKTIRYSGKRDATGLDYGYRYYPQWAGRWLSTD 651
Qy 652 PAGTIDGLNLYRMVRNPNVTLLDPGLMPT-----IAERIAALKK 691
Db 652 PAGTVDGLNLFMRVRNPNVTLLFDSNGRISTGQEARRLVGEAFVHPLHMPVFERISVERK 710

RESULT 15
Q66A99 YERPS
ID Q66A99 YERPS PRELIMINARY; PRT; 994 AA.
AC Q66A99
DT 25-OCT-2004 (TREMREL. 28, Created)
DT 25-OCT-2004 (TREMREL. 28, Last sequence update)
DE Putative insecticidal toxin complex.
GN OrderedLocusNames=YPTB2233;
OS Yersinia pseudotuberculosis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_taxID=633;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=IP32953 / Serotype I;
EX PubMed=15358858; DOI=10.1073/pnas.0404012101;
RA Chain P.S.G., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O.,
RA Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L.,
RA Brubaker R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C.,
RA Simonet M., Chenal-Francisque V., Souza B., Dacheux D., Elliott J.M.,
RA Derbise A., Hauser L.J., Garcia E.;
RT "Insights into the evolution of Yersinia pestis through whole-genome
RT comparison with Yersinia pseudotuberculosis.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:13826-13831(2004).
DR EMBL; BX936398; CAH21471.1; -; Genomic_DNA.
DR InterPro; IPR005165; Anthrax_toxA.
DR InterPro; IPR006530; YD.
DR Pfam; PF03497; Anthrax_toxA; 1.
DR Pfam; PF05593; RRS_repeat; 4.
DR TIGRFAMs; TIGR01643; YD_repeat_2x; 2.
KW Complete proteome.
SQ
SEQUENCE 994 AA; 109260 MW; 76C8CE064806F46F CRC64;

```

```

Query Match 43.4%; Score 2173; DB 2; Length 994;
Best Local Similarity 53.3%; Pred. No. 1.3e-107;
Matches 451; Conservative 105; Mismatches 208; Indels 82; Gaps 14;

Qy 8 LVQKPTVSVVDNRGLIIRNDFHRTTANGO-PDTRITRHQYDIHGLNQSIDPRLYEAK 66
Db 5 LFSKTPSVTVLDNRGLTVRDYAHRRPSPDVTSRITHQYDARGFLTQSDAPLHGAG 64

Qy 67 QTNNTIKPNFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLVTV-----AT 114
Db 65 LM-----NFSYLTDLTGRIILRTQAGNDGTTVSLNDATGRPFISASNISSDGDGTRGQ 118

Qy 115 GVITQRTQVETSLPGRLLSVAEQTEPEKTSRTERLIWAGNTEAEKOHNLACQCVRYHYDT 174
Db 119 AMTRTWQYEASLPGRVSVTEQV-TGKATRITERFVVAANTDAEKSINLAGACVSHYDT 177

```

```

Qy 175 AGVTRLBSLSLTGTVLSSQSLIDTQE-----ANWTGDNETVWQNLADDIYTTLTSTPDA 230
Db 178 AGLVQPSIALTGVPLSVTRILMKASADNPDAVADWQGDASAWNDQLDGETHTHTLTITADA 237
Qy 231 TGALLTOTDAKGNLQRLAYDVAGOLNGSWLTLKGTEQVILKSLTYSAGOKLREHGND 290
Db 238 TGAULTTTDAKGNLQRMAYDVAGLLSGSWLTLKGTEQVIVKSLTYSAGOKLREHGNG 297
Qy 291 VITEYSYEPETQRLIGIKTRRP-----SDTKVLODLRYEYDPVGNVISIRNDAEATRFWHN 346
Db 298 VVTIYEYEPETQRLVGIKTERPAGHASCAKVLQDLRYEYDPVGNVLKVTNDAEETFRWRN 357
Qy 347 QKVPENTYTVDSLYQLISATGREMANIQOSHQPSPA--LPSDNTYTNTRYTYTDR 404
Db 358 QKVPENTYTVDSLYQLVSATGREMANAQOQSCSLPSTTVFLPADSSAYTRYSTRYTYDE 417
Qy 405 CGNLTKLQHSPPATQNNYTNITVSNRNRVAVLSTLTEDPAQVDALFDAGGHONTLISQ 464
Db 418 AGNLTQIKHNPATNNSYTTIKLTVSDRNRGVSLTLENAAVDVALFTAGGQOTQLOPQ 477
Qy 465 NLNNTNRGELQOQVTLVKRDKGANDREWYRYSGGRRMLKINEQOASNAQRTVRYLPN 524
Db 478 HLIWTARNELLKVTPVRED-GSTDDSESRYDAASQRILKVSROKTNTSMQTRVLYLPG 536
Qy 525 LELRLTONSTATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQL 584
Db 537 LELRSKSGDTETEGLOVITVGEAGRAQVRVLHWSGKPEDIDDDQIRYSYDNLGSCSL 596
Qy 585 ELDSEGOIISSEYYPYGGTALWAARNQTEASYKTIRYSGKRDATGLYYGYRYYPQWI 644
Db 597 ELGGDGNIIISAEYYPYGGTAVVRAVEADYKTIRYSGKRDATGLYYGYRYYPQWA 656
Qy 645 GRWLSSDPAGTIDGLNLYRMVRNPNVTLLDPGLMPTIAERIAALKKQKVTDSAPSPANA 704
Db 657 GRWLSADPAGMVDGLNLFMRARNPNVAFIDRNLNSELAYS-QAFKRT-----ANK 706
Qy 705 TNVAINTRPPVAPKPSLPKASTSSQPTTHPIGAANIKP--TTSGSSIVAPLSPVG----- 757
Db 707 YNVIIGVRAP-----NPLGETLLKEGFPSPKPMKAKSSPTGTAGF 748
Qy 758 ---NKSTSEISLPESAQSSSSSTTSTNLQKSFITYRADNRSFE-----EMQ 801
Db 749 IAEPTIYSKVSFSAKKQORASIDKAKALGSESIDLFISKSRINELIDTGNLSLGENRYS 808
Qy 802 SKFPEG 807
Db 809 AKYPVG 814

```

Search completed: February 16, 2006, 21:43:00
Job time : 133.785 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	2537.5	50.7	915	6	US-10-647-956A-6	Sequence 6, Appli
2	183	3.7	2828	7	US-11-080-991-54	Sequence 54, Appl
3	183	3.7	2828	7	US-11-186-284-49	Sequence 49, Appl
4	175	3.5	2725	7	US-11-113-424-52	Sequence 52, Appl
5	174.5	3.5	1417	7	US-11-052-554A-8	Sequence 8, Appli
6	171.5	3.4	1345	7	US-11-052-554A-282	Sequence 282, Appl
7	164.5	3.4	5024	6	US-10-793-626-2964	Sequence 2964, Ap
8	160.5	3.2	1579	7	US-11-052-554A-9	Sequence 9, Appli
9	156	3.1	2314	7	US-11-013-759-11	Sequence 11, Appl
10	154.5	3.1	1468	6	US-10-467-657-1088	Sequence 1088, Ap
11	153.5	3.1	2769	7	US-11-113-424-14	Sequence 14, Appl
12	153	3.1	1045	7	US-11-113-424-54	Sequence 54, Appl
13	153	3.1	1094	6	US-10-821-234-1097	Sequence 1097, Ap
14	153	3.1	2376	7	US-11-096-051-4	Sequence 4, Appli
15	153	3.1	2715	7	US-11-096-051-2	Sequence 2, Appli
16	153	3.1	2715	7	US-11-113-424-51	Sequence 51, Appl
17	153	3.1	2721	7	US-11-096-051-10	Sequence 10, Appl
18	153	3.1	2725	7	US-11-096-051-8	Sequence 8, Appli
19	152.5	3.0	1562	7	US-11-052-554A-211	Sequence 211, Appl
20	152.5	3.0	2515	7	US-11-113-424-53	Sequence 53, Appl
21	151.5	3.0	2053	7	US-11-013-759-9	Sequence 9, Appli
22	151	3.0	5251	7	US-11-052-554A-281	Sequence 281, Appl
23	150.5	3.0	1992	7	US-11-013-759-3	Sequence 3, Appli
24	150.5	3.0	1932	7	US-11-013-759-13	Sequence 13, Appl
25	150.5	3.0	2047	7	US-11-013-759-4	Sequence 4, Appli

```
Qy 300 ETORLIGKTRPSDVKVLDRYEDPVGNVISIRNDAEATRFHKNKMPENTYTYDS 359
Db 293 ETORLIGITTRPSDAKVLDRYQYDPVGNVINIRNDAEATRFWRNOKVAPENSYTYDS 352
Qy 360 LYQLISATGRMANIGQOSHOPFALPSPDNNVTNTYTRTYDRGGLTKIQRHSSPATQ 419
Db 353 LYQLISATGRMANIGQONQLPALPSPDNNVTNTYTRSYDHSNGLTQIRHSSPATQ 412
Qy 420 NNYTNTITVNSRNLAVLSTLTPDAQVADLFDAGGHQNTLISGNLWNRGSLQVTL 479
Db 413 NNYTVAITLSNRNRGLSTLTPDNPQVDTLFDAGGHQNTLISGNLWNRGSLQVTL 470
Qy 480 VKRDKGANDREWYRSGDGRMLKINEQOASNAQTORVTLNLHLRLTONSTATED 539
Db 471 ---NNGPGN--EWYRDSNGRQLKVSQEQPTNTQOORVILYLPGLRLTQSNATTEE 525
Qy 540 LQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLDSEGOIISEBEYY 599
Db 526 LHVITLGEAGRAQVRVLHWESGKPEDVNNQLRYSYDNLIGSSQLDQNOQOIISEBEYY 585
Qy 600 PYGCTALWAARNQTEASYKTIRYSGKRDATGLYYGYRYYPQWIGRWLSSDPAGTIDGL 659
Db 586 PFGCTALWAANSQTEASYKTIRYSGKRDATGLYYGYRYYPQWAGRWLSADPAGTIDGL 645
Qy 660 NLYRMVRNVPVLLDPGLMPTIAERIAALKKVKVT 695
Db 646 NLYRMVRNVPVSLQDENG LAP-----EKGYT 672

RESULT 2
US-11-080-991-54
; Sequence 54, Application US/11080991
; Publication No. US20050266437A1
; GENERAL INFORMATION:
; APPLICANT: Veiby, Petter Ole
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR
; TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY OF BREAST
; TITLE OF INVENTION: AND OVARIAN CANCER
; FILE REFERENCE: MRI-039
; CURRENT APPLICATION NUMBER: US/11/080,991
; CURRENT FILING DATE: 2005-03-11
; PRIOR APPLICATION NUMBER: US/10/176,847
; PRIOR FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 112
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-080-991-54

Query Match 3.7%; Score 183; DB 7; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.0017;
Matches 165; Conservative 108; Mismatches 276; Indels 302; Gaps 38;

Qy 307 IKTRPSDVKVLQ---DLRYEDPV-----GNVLSI-RNDAEATRFHNNQ 347
Db 542 IKSMEPSDGLYQCTAQVRDEMDRVRVLVQSPSTQPAEKDVTIGKNPGESVTLPCNA 601
Qy 348 KVMPENTYTVDSLVQLISATGREMANIGQOSHQPSP-----ALPSDNNVTNTYTRTYD 403
Db 602 LAIPEAHLW-----ILPNRRIINDLANTSHVWMLPNTGISPKVQVSDSGYTRCVAVN 655
Qy 404 RGGNLTQIOWSSPATONNTYTNITVS-----NRSNRAVLSTLTPDAQVADLFD 452
Db 656 QQG-----ADHFTVGIITVTKGSLPSKRGRRPCAKALSVRDIVE-----D 698
Qy 453 AGGHQNTLISQNLWNRGEL-----QQVTLVKRDKGANDREWYRSGDGRMLKI--- 505
Db 699 EGG-----SGMGDEENTSRLLHPKQOEVLTKDKDAINGDKK-----AKGRRLKLWK 748
Qy 506 --NEQOASNAQTORV-----T 520
```

```
Db 749 HSEKEPTNVAEGRVFESERRINMANKQINPERWADILAKVRGNLPGKTEVPPLIKTT 808
Qy 521 YLPNLELRLLTON-----STATTEDLOVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYD 576
Db 809 SPPSLSLLEVTPPPFPAVSPSASPVQTVTSAEESADVPLLEE-----E 852
Qy 577 NLIG-----SQLELDSGOIISEEYYPYGGTALWAARNQTEASYKTIRYSGKRDATG 631
Db 853 HVLGTIISASGLENHNGVILVEPV-----TSTPLEEVVDDLSKTEETITSTEGDLKG 907
Qy 632 -----LYYGYRYQP-----WIGRWLSSDPAGT-----IDGL 659
Db 908 TAAPTLISEPEPSPVTLTLDVYKETHETATEGWSAADVGSSPEPTSEYEPPLDAV 967
Qy 660 NLYRMVRNVPVLLDPGLMPTIAERIAALKKVKVTSAPSP-----ANATVAINIRPPVA 716
Db 968 SL---AASEPQVFPD-D-LETKSQPDDEKMKEDTFAHLTPTPTTIWVNDSSSTQLFEDSTI 1023
Qy 717 PKPSLP-----KASTSQPT-----THPIGAANI---KPTTSGSS 748
Db 1024 GEPGVPQSHLQGLTDNIHLVKSLSLTDILLIKKMKEMSQTLOGGNMLEGDPHRSRS 1083
Qy 749 -----IVAPLSPVG-----NKSTSEISLFPESAQSSSSSTTS 779
Db 1084 ESEQESKSLTLPDSTLIGIMSSKSPVKKPAETTVGTLLDKDQTTVTTPRKVAPSSTMS 1143
Qy 780 TNLQKSFLLYRADNRSFEBEQSKPFGFKAWPLDPTMARQFASIFIGKQDNTNLPKET 839
Db 1144 THPSR-----RRPNRRRLRPKNFRHRKQTPP-----TTFAPSET 1179
Qy 840 VKNISTWGAPEKPKOLSNYIKYTKDKSTVWVSTAINT-----EAGGSSGAP----- 886
Db 1180 FSTQPT--QAPDIK-ISSQVE-SSLVPTAWVDNTVNTPKOLEMKNAPETS KGP RKHG 1235
Qy 887 ----LHKIDMDLYEFAIDGQKLNPLPGRGTNNMVPFLLDTPQIET-----SSIIALNHGP 938
Db 1236 KRPKNKHYTSTVSSRASGSKSPSPENKHNIV-----TPSSETILLPRVSLKTEGP 1289
Qy 939 VNDABISFLTT 949
Db 1290 YD--SLDYMTT 1298

RESULT 3
US-11-186-284-49
; Sequence 49, Application US/11186284
; Publication No. US20050266493A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Berger, Allison
; APPLICANT: Guillemette, Tracy L.
; APPLICANT: Kamathkar, Shubhangi
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Thibodeau, Stephen N.
; APPLICANT: Burgart, Lawrence J.
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF COLON CANCER
; FILE REFERENCE: MPM01-029P2RNM
; CURRENT APPLICATION NUMBER: US/11/186,284
; CURRENT FILING DATE: 2005-07-21
; PRIOR APPLICATION NUMBER: US/10/301,822
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 60/339,971
; PRIOR FILING DATE: 2001-12-10
; PRIOR APPLICATION NUMBER: US 60/361,978
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/381,988
; PRIOR FILING DATE: 2002-05-20
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: FastSeq for Windows Version 4.0
```

```

; SEQ ID NO 49
; LENGTH: 2828
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-186-284-49

Query Match      3.7%; Score 183; DB 7; Length 2828;
Best Local Similarity 19.4%; Pred. No. 0.0017;
Matches 165; Conservative 108; Mismatches 276; Indels 302; Gaps 38;

QY 307 IKTRPSPTKVLQ-----DLREYEDPV-----GNVISI-RNDAEATRFWNO 347
DB 542 IKSMFSDGLGQCIQAQVDEMDRVMYRVVLSVSPSTQPAEKDTVTIGKNGPESVTLPCNA 601
QY 348 KMPENTTYDLYQLISATGREMANIGQSHQFPSP-----ALPSDNNNTYTYTYVD 403
DB 602 LAIPEAHLW-----ILPNRIINDLANTSHVYMLPNGTSLIPKVOVSDSGYRCVAVN 655
QY 404 RGNLTKIQHSSPATONNTYTVS-----NRSNRAVLSTLTEDPAQVDALFD 452
DB 656 QQG-----ADHFTVGITVTKGSLGSLPSKRGRRPGAKALSRRVEDIVE-----D 698
QY 453 AGGHQNTLISGQNLNWRNGEL-----QQVTLVKDKGANDREWVRYSGDGRMLKI-- 505
DB 699 EGG-----SGMDENTSRRLHPKDEVLTKDKDAINGDK-----AKGRRKLKLWK 748
QY 506 --NEQASNNACTQRV-----T 520
DB 749 HSEKEPTNVAEGRRVFESRRRINMANKQINPERWADILAKYRGKLPKGTVPPLIKTT 808
QY 521 YLPNLELTKQN-----STATTEDLQVITVGEAGRAQVRVLHWSGKPEDIDNNQRYSD 576
DB 809 SPPSLSEVTPPPAVSPSPASPVQTVTSAESSADVPILGEE-----E 852
QY 577 NLIG-----SSOLEDSQIISSEBYYPYCGTALWAARNQTEASVKTIRYSGKERDATG 631
DB 853 HVIGTISSASMGLEHNGHVVILVEPV-----TSTPLEVVDLSEKTEBITSTEGDLKG 907
QY 632 -----LYYGYRYYP-----WIGRWLSSDPAGT-----IDGL 659
DB 908 TAAPTLISEPYEPTLHTLDVTYKEPTHEETATEGMSAADVGSSPEPTSSSEYEPPLDAV 967
QY 660 NLYMRVNNPVTLLPDGLMPTIAERIAALKKQKVTDSAPSP-----ANATNVAINRPPVA 716
DB 968 SL---AESEPMQYFPD-LETQSDQEDKMKEDTFAHLTPTPTIWNDSSTSQLFEDSTI 1023
QY 717 KPPLP-----KASTSQPT-----THPIGAANI---KPTTSGSS 748
DB 1024 GEPGVPGOSHLOGLTNTLHLVKSLSUSTQDTLLIKGKMKEMSTQLQGNMLEGDPHRSKS 1083
QY 749 -----IVAPLSPVG-----NKSTSRIISLPESAQSSSSTTS 779
DB 1084 ESEGESKSIITLPDSTLIGMSSSPVKKPAETTVGTLDDKDTTITVTPRKVAPSSTMS 1143
QY 780 TNLQKKSFTLVADNRSEEMQSKPEPFKAWTPLDTKMAQFASIFIGQKDTSNLPKET 839
DB 1144 THPSR-----RRPNRGRRLRNKFRHRKQTPP-----TTPAPSET 1179
QY 840 VKNISTGAKPKLKDLSNVIKYTKDKSTVMYSTAINT-----EAGQSSGAP----- 886
DB 1180 FSTQPT--QAPDK--ISSQVE--SSLVPTAWVDNTVTPKQLEMEKNAEPTSGTPRKHG 1235
QY 887 ----LHKIDMDLYEPAIDQKLNPLPEGRTKMVPSSLTDTPOIET-----SSIIALNHPG 938
DB 1236 KPNKHYRTPPSTVSSRASGSKSPSPENKRNIV-----TPSSETILLPRVSLKTEGP 1289
QY 939 VNDABISFLT 949
DB 1290 YD--SLDYMTT 1298

RESULT 4
US-11-113-424-52

```

```

; Sequence 52, Application US/11113424
; Publication No. US20050260713A1
; GENERAL INFORMATION:
; APPLICANT: Gangolli et al.
; TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-225
; CURRENT APPLICATION NUMBER: US/11/113,424
; CURRENT FILING DATE: 2005-04-21
; PRIOR APPLICATION NUMBER: 60/256,704
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/311,590
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/257,314
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/311,613
; PRIOR FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: 60/315,617
; PRIOR FILING DATE: 2001-08-29
; PRIOR APPLICATION NUMBER: 60/307,506
; PRIOR FILING DATE: 2001-07-24
; PRIOR APPLICATION NUMBER: 60/322,358
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/294,075
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: 60/288,153
; PRIOR FILING DATE: 2001-05-02
; NUMBER OF SEQ ID NOS: 190
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 52
; LENGTH: 2725
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-113-424-52

Query Match      3.5%; Score 175; DB 7; Length 2725;
Best Local Similarity 22.3%; Pred. No. 0.0053;
Matches 173; Conservative 85; Mismatches 256; Indels 262; Gaps 44;

QY 18 YDNRLGIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDPRLYEAKQTNTTIKPNFL 77
DB 1865 YDQSGKII-----SRTWADG---KIMSYTY-----LEKSVMLLLHSQRR-----YI 1902
QY 78 WOYDLTGPNLCTESIDAGRTVTLNDIEORPLLTVTATGVIQ-----TRQYETSSLP 128
DB 1903 FEYD---QPDCLLS-----VTPESMVRHSLQTMLSVGYRNIYTPPDSSTSFIDYSRD 1953
QY 129 GRLLSVASQTPPEKTSRITERLIWAGNTEAEKHNLACQVRHYDTAGVTRL--ESLSLT 186
DB 1954 GRLL-----QTLHLGTGR---RVLYKYTKQARLSEVL-----YDTQTTLTYESSGVI 1999
QY 187 GTVLSSQSQLIID---TOEANWTGD-----NETVQNMALADDIYTTLTSTFDATG- 232
DB 2000 KTI-----HLMHDPICITRYRQTGLPIGRQIFRESEGLVNARFDYSY---NNFRVTSM 2051
QY 233 -ALLTQTDAKNIQRLAYDVAGQLN--GSWLTGKQQTQVILKSL-----TVSAAGOKLR 284
DB 2052 QAVINETPLPLDLRYV-DVSGRTEQFGKFSVINDLNOVITTTVMKHTKIFSANGQ--- 2107
QY 285 BEHGNVDITEYSYRPETQRLIGIKTRRPSDTKVLQDLRY---EYDPVGNV---ISIRN 336
DB 2108 -----VIEQVQE-----ILKAIAYWMTIQDNGVRMVICDIRVGV 2142
QY 337 DAEATRFWHNOKVMPENTTYDLYQLISATGREMANIGQSHQSPSPALPSDNNNTYNY 396
DB 2143 DANITRYPYE-----YDADGQLQTVSV-----NDKTCQW 2170
QY 397 TRTTYDRGGLTKIQHSSPATONNTYTNITVNSRNPRAVLSTLTEDPAQVDALFDAGGH 456
DB 2171 --RYSYDLNGNINLLSHGKSA-----RLTPLRYDLDRITRLGEIQYKMDB--DGFILR 2219
QY 457 QNTLISGQNL--NWNTRGELQQVTLVKRDKGANDREW---YRSGDGRMLKINKINQQASN 512
DB 2220 QR-----GNDIPEYNSNGLQ-----KAYNKASGWTVQYYDGLGRRV-----ASK 2260

```

```

Qy 513 NAQTQ-----RVYLPNLELRLTONSTATTEDLOVITVGEAGRAQVRVLHWE 559
Db 2261 SSLGCHLQFFVADTNPIRVTHLYN---HTSSEITSLYYDLO-----GHLIAMELS 2308
Qy 560 SKPEDIDNNOLRYSYDNLIGSSQLELDSEGOITSEBEYYPYGGTALWAARNOTEASYKT 619
Db 2309 SGE-----EYVACDN-TGTPLAVFSSRGQVKEILYTPGD-----IYHDTYPPQV 2355
Qy 620 I--RYSGKERDATGLYYGYRYQFWIGRW-----LSSDPAGTIDGLNLYRMV 665
Db 2356 IIGPHGLYDFLTKLVHGLQRDYDVVAGRWTPPNHHIWKQLNLLPKP-----ENLYSFE 2409
Qy 666 RNNPVTLDDPCLMPTTAERTAAALKKQVTDASAPANATVAINIRPPVAPKSL 721
Db 2410 NNYFV-----GKIQDVA-----KYTTDIRSWLELFGQLHNVLPGPPKPEL 2450

RESULT 5
US-11-052-554A-8
; Sequence 8, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 8
; LENGTH: 1417
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
; US-11-052-554A-8

Query Match 3.5%; Score 174.5; DB 7; Length 1417;
Best Local Similarity 19.3%; Pred. No. 0.0022;
Matches 209; Conservative 122; Mismatches 327; Indels 427; Gaps 52;

Qy 107 PLLTWTATGVTQRYETSSLPGRLLSVAEQTPBEK---TSRITERLIWAGNTE--AEKD 161
Db 291 PLLTLSA-GHKQKSGENDTRFGLVNYRIGELPLAKQLDTSIRERRVLGSRVLDLVERN 349
Qy 162 HNLAQCVRHVDYTAGVTEL-----EELSITGTVLSQSSQLLIDTQEAHW----- 205
Db 350 NNI-----VLEYRKSEVIRIALPERIEGKGOTLSL-GLVSKATHGLKNVQ---WEAPSL 401
Qy 206 -----TGDNETVQNNML-----ADDTYTLSTFDATGALLTQTDAGNI-QRLAYDV- 251
Db 402 LAEGKLTGQG-SQWQVTLPAYRCKQNYAISA-----VAVDNKGNTSKRVQTEV 452
Qy 252 ---AG-QLNGSWLTLKGQTE-QVTKISLTYSAAQKRLREHGNVDVITEYSVEPTQRLIG 306
Db 453 ITGAGMSADRTALTLDGQSRIQMLANGNEQPLVLSLRDAEG-----QPVTGMKDQ 503
Qy 307 IKTRRPSTDKVLQDLRYEYDVGNVISIRNDAEATRFWNNKVPENTYTYDLSYQLISA 366
Db 504 IKT-----ELTFKAGNIVT---RSLKATKS-QAKPTLGEFTETEAQVQSFT 548
Qy 367 TG-----REMANI-----GOOSHQFFSPA 385
Db 549 TGTQSGERATITVSDGMSKVTYAEIRATIMDVANSTLSANPESGDVADGQQAATLTLTA 608
Qy 386 LPSDNNVTYN-----YTRYTYDRGNLT----- 409
Db 609 VDSEGNPVTGEASRLRFVPODTNGVTGVAISEIKFPGVYSAAVSVSTRAGNVVVRFAFSEYQ 668

```

```

Qy 410 -----KIOHSS-----PATQNNYTTNIITVSNRSNRAVLSTLTEDPAQV 447
Db 669 LGTLOQTILKFVAGPLDAAHSITILNPKPVPVGGTTVAITWTVKDAYDNPVSLTPEAPSLA 728
Qy 448 DALFDAGGHQNTLISGQNLWNTRGE---LQQVTLVKRDKGANDDREYRYSGDGRMLK 504
Db 729 GA-----AAGGSTASG---WTNNGDGTWTAQITL-----GST-----AGELEVMPK 766
Qy 505 INEQOASNNQAQRTVYLPNLELRLTONSTATTEDLOVITVGEAGRAQVRVLHWSGKPE 564
Db 767 LINGQAAANA--AKVTVAD--ALSSNOSKVSVAED-----HYKAGEST 806
Qy 565 DIDNNQLRYSYDNLIGSSQLELDSEGOITSEBEYYPYGGTALWAARNOTEASYKTIRYSG 624
Db 807 TV-----TLVAKDAHGNAIS-----GLALSASLTGTASEGATV--SS 841
Qy 625 KERDATGLYYGYRYQFWIGRWLSSDPAGTIDGLNLYRMVNNPVTLLDDPDGLMPTIAE 684
Db 842 WTEKNGSY-----VATLTGKGTGELRVMLPF-----NGQPAATEAAQLTVIAG 886
Qy 685 RIAALKKQKVTDS-APSPANATVAINIR-----PPVAPKPSLP-----KASTSQ----- 729
Db 887 EMSSANSTLVADNKAPTVTKTTELTFTVKDAYGNPVTGLKPDAPVFGAASSTGSRPSAG 946
Qy 730 -----PTTHPIGAA---NIKPTTSGSSIVAP---LSPVGNKSTSEI----- 764
Db 947 NWTGKNGVYVSTLTLSAAGQLSVMPRVNGQNAVAQPLVLNVAGDASKAIEIRDWTVKVN 1006
Qy 765 -----SLPESAQSSSSSTSTN----- 781
Db 1007 NQLANGOSANQITLTVVDTYGNPLQGGQEVTLTLPGQVTSKTGNTVTNTNAAGKADIELMST 1066
Qy 782 -----LOKESFTLYRAD-----NRSFEEMOSKFPFGKAWTPLDTKMARQF 822
Db 1067 VAGEHNISASVNGAOKTVTYVKFNADASTGQANLQVDAQAQVANGKDAFT-LTANVEDKN 1125
Qy 823 ASIFIGQKDTSNLPKETVKNISTWGAQPKLKLDSNYIKYTKDKSTVWVSTAINTEAGGQS 882
Db 1126 GNPVPGSLVTFNLP-----GVKPLTCD-----NVWV----- 1152
Qy 883 SGAPLHKIDMDLYEFAIDGQKLNPLPEGRTKNNMVPSSLDDTPQIETSSIIALNHGPVND 942
Db 1153 -----KAN--DEGKAEQLQVSVTAGTYEITAS---AGNSQPSNTQ 1187
Qy 943 EISPL 947
Db 1188 TITFV 1192

RESULT 6
US-11-052-554A-282
; Sequence 282, Application US/11052554A
; Publication No. US2005028866A1
; GENERAL INFORMATION:
; APPLICANT: Sachdeva, et al.
; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
; TITLE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
; FILE REFERENCE: 30853/40359A
; CURRENT APPLICATION NUMBER: US/11/052,554A
; PRIOR FILING DATE: 2005-02-07
; PRIOR APPLICATION NUMBER: US 60/589,227
; PRIOR FILING DATE: 2004-07-20
; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
; PRIOR FILING DATE: 2004-02-06
; NUMBER OF SEQ ID NOS: 763
; SOFTWARE: Patent in version 3.3
; SEQ ID NO 282
; LENGTH: 1345
; TYPE: PRT
; ORGANISM: Escherichia coli 0157:H7
; US-11-052-554A-282

Query Match 3.4%; Score 171.5; DB 7; Length 1345;

```

Best Local Similarity 19.4%; Pred. No. 0.0032;
Matches 219; Conservative 140; Mismatches 367; Indels 401; Gaps 55;

QY 3 NIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDPRL 62
DB 14 NNDTKVNIAPAS-----NAQVTINIPAOVVVNNSDSVQLTATVKDPSNHPVAGI----- 65

QY 63 YEAKQTNNTIKPNFLWQYDLTGCPCTESIDAGRTVTTLNDIEGRPLLTATGVIQTRQY 122
DB 66 ----TVNFTMPQDVAAANFTLENNGIAITQANGEAHVTL---KGKAGTHVTATLSNNNT 118

QY 123 ETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKHNLACQCVRYHDYAGVTRLES 182
DB 119 SDSQ---PVTFFVADKTSALVQLQISKNEI-TGN-----GV---DS 151

QY 183 LSLGTVLSSQSOLLIDTQEANWTGDNETVQNMMLADDIYTLSTFDATGALLTQTDAGK 242
DB 152 ATLTAIVKQDF-----DNE-----VNNLPVTFST-ASSGLTLFPGESNT 189

QY 243 NIQRLAYDVAGOLGSMWTLKQGT--EQVIKSLTYSAAQOKLREHGNVDITEYSYPE 300
DB 190 NESGIAQ-----ATLAGVAFGEQTVTASLANNGAS-----DNK 222

QY 301 TORLIGIKTRRPSDKVLQDLRYEYDPV-----GNVISIRNDAEATRFWNOQ 347
DB 223 TVHFIG-----DTAAAKII--ELTPVPDSIIAGTPQNSSGSVIT-----ATVVDNNG 267

QY 348 KVPENTYVDSLYQLIGATGREMANIGQSHQFPSPALPSDNN---TYTNTYTRTYVD 403
DB 268 FPKGVGVVNFYS-----NAAATMNGGO-----AVTNEQKATVITYN-TRS-SIE 312

QY 404 RGGNLTKIHSSPATQNNYTNITVSNRSRAVLSTLTPEDPAQVDALFD---AGGHQNTL 460
DB 313 SGARPDVTEASLENGSSTLSINUNADASTAHLTLL-----QALFDTVSAGDTNLY 365

QY 461 ISGQNLNWNTRGELQOVTL-VKRDIG-----ANDREWY-RYSGDGRMLKINEQ 508
DB 366 LEVKD-NYGNVGPQOEVTLSPSPGVTTPSNNAIYTNHNDGNFYASFATKAGVYQVAT 424

QY 509 QASNAQQRVTYLPNL---ELRL-----TONSTATT----- 537
DB 425 LENGDSMQOVTYVFNVAEISLAASKDPVIANNDITLTATVADTEGNAIANSEVTF 484

QY 538 ---EDLQV-ITVGEAGR-----AQRVVLHBSGKPEIDNN----- 569
DB 485 TLPEDEVANFTLDGCKVVDTEGKAVTLKGTAKAGHTVVTASMAGGSEQLVWNFIADT 544

QY 570 -----OLRYSVDNLI-----GSSQLE---LDSEGQIISBEEY-----YPYGGT 604
DB 545 LTAQVNLNVTEDNFLANNVGMTRQLQATVTDGNGNPLANEAVTFTLPADVSAFSLTGQGS 604

QY 605 ALWAARNQTEASYKTIIRSGKRDATGLYYGYRYQPWIGWLSDDPAGTID---GLNL 661
DB 605 AITDINGKAEVT-----LSGTSKGYTPVTVSVNN 633

QY 662 YRMVRNPNVTLDPDGLMP-----TIAERIAALKKKNKVTDSAPSANATNVAINI 711
DB 634 YGVSTKQVTLIADAGTAKLASLTSVSVFVSTTEGATWTASVTDANGNPVE--GIKNPF 691

QY 712 RPPVAPKPSLPKASTSSQPTTHPIGAANIKFTTSGSSIVAPLSPVGNKSTSEISLPESAQ 771
DB 692 R-----GTSVTLSSTSVE--TDDRGAELIIVTSTEVGLTKVSASLADKPTVISRLNAK 744

QY 772 SSSSTSTTNLKKSFLLYRADNRSEFMQSKPEG----- 807
DB 745 ADINSAITISLE-----IPEGQVNVAAQDVAAHVNDQFQNPILN 784

QY 808 ----FKAWTFLDTKMAROPASIFIGKQTSNLPKKTV---KNISTWGAQPKLKLDSNYIK 860
DB 785 BSVTFSAEPPEHMTISQNIIVS-----TDTHGIAEVTWTPERN-----GSVMVKASLANGSS 835

QY 861 YTKD-----KSTVWVST---AINTAEGG-----QSSGAPLHKIDMDLYEFAIDGQKL 904
DB 904 YTKD-----KSTVWVST---AINTAEGG-----QSSGAPLHKIDMDLYEFAIDGQKL 904

DB 836 YEKDLVIDQKLTLSASSPLGVNSPTGATLTATLTSANGTF-----VEGQVI 883

QY 905 N--PLPEGR-----KMWPSLLDTPQIETSSIIALNHGPV 939
DB 884 NFSVTPGATLSGGKVRNTSSGQAP-VVLTSNKVGTVTVTASFHNGV 929

RESULT 7
US-10-793-626-2964
; Sequence 2964, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: P034800S
; CURRENT APPLICATION NUMBER: US/10793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn.Ver. 2.1
; SEQ ID NO 2964
; LENGTH: 5024
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
; FEATURE:
; NAME/KEY: MOD_RES
; LOCATION: (5024)
; OTHER INFORMATION: variable amino acid
US-10-793-626-2964

Query Match 3.3%; Score 164.5; DB 6; Length 5024;
Best Local Similarity 20.4%; Pred. No. 0.064;
Matches 208; Conservative 158; Mismatches 407; Indels 249; Gaps 51;

QY 8 LYOKTPTV---SVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDPRLYE 64
DB 7 LVAKQPTVKQTSVINE-----DQEQSAYND---SITMGQTIINKTADPVLDTKLVLD 56

QY 65 AKQTNVTKPNFL-WQYDLTGCPCTESIDAGRTVT--LNDIEGRPLLTATGVIQOT--- 119
DB 57 NAINISITKENALGEQKLT--TAKTEAINALNTLADLNTPKQEAIKTAINTAHTRTDVT 114

QY 120 -ROYETSLSPLGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDH-----NLAGOCVRHY 172
DB 115 AEQSKANQINSAMHTLRQNISDSNVNESVI---NAEPEKQHAFTALNNAKEIVNE- 170

QY 173 DTAGVTRLESLSLTGTVLSSQSOLLIDTQEANWTGDNETVQNMMLADDIYTLSTFDATG 232
DB 171 -----QATLDANSINQAKALLITKNA-LDGEQLRRAKENADQEIINTLNLQ--- 217

QY 233 ALLTQTDAGNIQRLAYDVAGOLGSMWTLKGTQEQVIKSLTYSAAQOKLREHGNVDI 292
DB 218 ----TDAQRNSEK-----GLVNSS-----QTRTEVASQLAKAKELNKVMEQLNN--- 257

QY 293 TEYSVEPTQGLIGIKTRRPSDKVLQDLRYEYDPVGNVISIRNDAEATRFWNNOKVMPE 352
DB 258 -----LINGKQMINSSKPFINEDANQOQOAYSNAIA---SAEVLK---NKSQNP 300

QY 353 NTYTVDSLYQLISATGREMANIGQSH---QFPSPALPSDNNT---YTNTRT----- 399
DB 301 LDKV--TIEQAINNINSAINNLNGEAKLTAKEDAVASINNLISGLTNEQTKENQAVNGS 358

QY 400 YTYDRGNLTKIQHSSSPATQNNYTNITVSNRSRAVLST---LTEDPAQVDALPDAGCH 456
DB 359 QTRDQVANV--LRDSKALDQSMQTLRLDLVNNQN--VIHSTSNYFNESTQKNTYDANIDN 414

QY 457 QNTLISGON---LWNTEGE-LQQVTLVKRD-KGANDDREWYRYSGDGRMLKINEQOAS 511
DB 415 GSTVITGQHSBELNKSTIDQTSISQINTAKNDLHGAE-----KUQRDKGT 458

QY 512 NNAQTVRYLPNLELRTQSTATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQL 571
 Db ANEITGQLGYL-----NDPOKASBESLVN-GSNTSRVBE-HLNEAKSL---NNAM 504
 QY 572 RYSVDNLGSSOLELDSGQIISBEEYYPYGGTALWAARN-QTEASYKTIYRSGKERDAT 630
 Db 505 KQLRDKVAEKTWQSSDIYNDSTEHQRYG-DQALQEAENLINEIGNFTLNKSEIBQKLO 563
 QY 631 GLYYGYRYQFWIGRWLSSDPAGTIDGLNLYRMVRNPNVTLDPDGLMPTIAERIALK 690
 Db 564 QL-----TDAQNALQSHLLEBAKNAITEI-----589
 QY 691 KKVYTDSPANATNVAINTR-----PPVAPKPSLPK-ASTSSQPTTHPIGAAN-----739
 Db 590 -NKLT--ALNDAORQKAIENVOAQTIPAVNQOULTDREINTAMQALRDKVQGNVHQ 646
 QY 740 -----IKPTTSSSVIAPL-SPVGNKSTSEISLPESAQSSSSTSTNL--782
 Db 647 SNYFNEDEQPKHNDNSVQAGQTIIDKLQDIPMKNEIEQAL-----NQINTQTALSG 700
 QY 783 QKKSFTLYRANRSEFEWQSKPEGFKAWTPL-----DTKMARQFASIFIGQKDTGNLP 836
 Db 701 ENKLHTDQESTNRQIEGLSSLNTAQINAEKDLVNOQAKTRTDVAKLATAKEINSAMN- 759
 QY 837 KETVKNISTWAKPKLDLSNVIKYTKDKSTVM-----VSTAINTEAGQSSGAP 886
 Db 760 RDGIQN-----KEDIKRSSAYINADPTKVAYDQALQNAENINATPNVELNKATIEQA 813
 QY 887 LHKI-----DMD-LYEFPAIDGQK-----LNPLPEGRTKMWPSSLIDTPQIETSSIIA 933
 Db 814 LSRVQQAQDLGCVQOLANAKQAQTQTVNGLSLNDGQKREL--NLLINSANTRYKQEE 871
 QY 934 LN 935
 Db 872 LN 873

RESULT 8

US-11-052-554A-9
 ; Sequence 9, Application US/11052554A
 ; Publication No. US2005028866A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Sachdeva, et al.
 ; TITLE OF INVENTION: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE
 ; FILE OF INVENTION: PROTEINS OF THERAPEUTIC POTENTIAL
 ; FILE REFERENCE: 30853/40359A
 ; CURRENT APPLICATION NUMBER: US/11/052,554A
 ; PRIORITY FILING DATE: 2005-02-07
 ; PRIOR APPLICATION NUMBER: US 60/589,227
 ; PRIOR FILING DATE: 2004-07-20
 ; PRIOR APPLICATION NUMBER: IN 173/DEL/2004
 ; PRIOR FILING DATE: 2004-02-06
 ; NUMBER OF SEQ ID NOS: 763
 ; SOFTWARE: PatentIn version 3.3
 ; SEQ ID NO 9
 ; LENGTH: 1579
 ; TYPE: PRT
 ; ORGANISM: Escherichia coli 0157:H7
 US-11-052-554A-9

Query Match 3.2%; Score 160.5; DB 7; Length 1579;

Best Local Similarity 17.8%; Pred. No. 0.021; Matches 177; Conservative 151; Mismatches 374; Indels 295; Gaps 46;

QY 58 IDPLRYEAKQNTNTPK---NFLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLITVTAT 114
 Db 273 LDP-----VKGSGEIKPLVSSIQYKALGYNEAALEA-----AG 310
 QY 115 QVITQRYVE-TSSLPGRLLSVAOTPEKTSRITERLIWAGNTBAEKDHNLAGQCVRHVD 173
 Db 311 GKVSTSGKDIWTPLPGVRYFT---NTPETDNT-----WSIDVTAE---DVKGNSLRHEQ 357

QY 174 TAGVTRLESLS-----LGTGLVSSQSLIID 199
 Db 358 SMVVIQAPTLISQKDSLLSVNPLTVAAKKSTTTLVTAHSDGTTPVGLALQITSEGVQD 417
 QY 200 TOEANWTGDNETWQNMNLAADDIYTLTSTFDATGALLTQTDAGNIQRIQLAYDVAGQLGWS 259
 Db 418 ITLSDWT-DN-----GDGSYTIQTLTAGTTSQVTLTP-----QING--452
 QY 260 LTLKGQTEQVIIKSLTYSAGOKLREHGNVDITEYSYEBETQRLIGIKTRRPSDTKVLQ 319
 Db 453 -----ESAVKESIVNIVPVVSSRDHSITIDNVSYAGDDIKRVVELKDDSN-----Q 501
 QY 320 DLRYEYDPVGNVISIRNDAB-ATREFWKNOK--VMPENTYTY-----357
 Db 502 PVAYQKEELKAVTVENSKFGATIVHEBQPVVAANYPAYKQGTALRAQLSLHNWAPL 561
 QY 358 -DSLYQL-----ISATGREMANIQQSHQPPSPALPSDNNTYNTYRTYTYDRG 405
 Db 562 QSHIYNIEMQNKAQVATLSATNNDVADKKTFTLTINVDESNDPLTNHQVTFKNEKG 621
 QY 406 G-----NLTKIQHSPATQNNYTNITVSNRSLRAVLSTLTEDP-----444
 Db 622 SAEFVEPQONTDAYVATINMVSVQAEEN-TISATLPNGFSORIIAKFVSDSSTPKFKQ 680
 QY 445 --AQVDALFDAGGHONTLS-----GQNLNNTREGELQQVTLVK 481
 Db 681 LVADPDTIIAGNSQGSTLTALITDFHNPPLKDMKNFVAPGGSOLD-NYATTTDQSGIVR 739
 QY 482 ----RDKGANDREWYRYSYSGDGRMLKINEQQASNNAAQTQRTVYLPNLEL-LFONS---533
 Db 740 VHLTSSKAGS-----YSVDA--SLEVDK---NIHQSVTITVVPNREQSVMTLNAGSG 786
 QY 534 TATTEDLOVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNL-IGSSOLELDSGQI 592
 Db 787 SAIANNTNIVTL-----TASVKDVYGHPLPDEDVKFTLPASMTGNFTLSSETARTDANGDA 842
 QY 593 ISBEEYYPYGGTALWA--ARNOEASKYTIYRSGKERDATGLYYGYRYQFWIGRWLSS 650
 Db 843 VTLRGTKAGEFTVATLTRNTVA-YQVTFIGDTNSA-----QLQP-LTASLNS 891
 QY 651 DPAGTIDGLNLYRMVRNPNVTLDPDGLMPTIAERIALKKNKVTDSAPSPANAT--NVA 708
 Db 892 IVAGNSTGSTL-----TATIL--DAYQNPDKQLVTFQSNVDVLTSETETVNTLQOAT 942
 QY 709 INTRPPVAPKPSLPKASTSQPTTHPIGAANIKPTTSGSSIVA-----PLSPVGNKSTSE 763
 Db 943 VTMTSNIAGQHNV-VWSRKAQAQSDNKTFSLSVLPPDESSAKVISITGAEKTIIVGENITLR 1001
 QY 764 ISLPES-----AQSSTSTSTNLQKSFPLYRADNR--SFEEMQSKFPEGFKAWT 812
 Db 1002 ILVQDAFNWVIAGQVRVLSAQPTNITIGDTAYT-----DNGVAYVNLSTQPGVYQVTA 1057
 QY 813 PLDTKMARQF-ASIFIGQKD-TSNLPKETVKN-----ISTWAK 849
 Db 1058 TLDNNSSSKVDNVVANGKLELTSSKPTTVHNSGIIITATARNARGELMPGQIITFSVT 1117
 QY 850 PKLXDLN-----YIKTKDKSTVWSTAI 874
 Db 1118 PEGATLSNTGCVLTDQSGQAKVTILTSDKNVYTVTAI 1154

RESULT 9

US-11-013-759-11
 ; Sequence 11, Application US/11013759
 ; Publication No. US20050249747A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Loosmore, Sheena M.
 ; APPLICANT: Sasaki, Ken
 ; APPLICANT: Yang, Yan Ping
 ; APPLICANT: Klein, Michel H.
 ; TITLE OF INVENTION: RECOMBINANT HIGH MOLECULAR WEIGHT MAJOR OUTER MEMBRANE
 ; TITLE OF INVENTION: PROTEIN OF MORAXELLA
 ; FILE REFERENCE: 1038-921MIS:jb


```

; CURRENT APPLICATION NUMBER: US/11/013,759
; CURRENT FILING DATE: 2004-12-16
; PRIOR APPLICATION NUMBER: US/09/361,619
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-11-013-759-11

Query Match      3.1%; Score 156; DB 7; Length 2314;
Best Local Similarity 19.1%; Pred. No. 0.072;
Matches 222; Conservative 157; Mismatches 398; Indels 384; Gaps 59;

Qy 3 NIDPKLYQKTPVSVDNRG---LIIIRIDFHRTTANG-----DPDRITRHQYDI 50
Db 814 NVDEK-----TIEUTGNGTKNGIKGVKTTTLTTTANGKATNFSTTNDALV--NAXDI 865

Qy 51 HGHNLQSDIPRLYEAKQTNNTIKPNFLWQYD-LTGNPLCTRSIDA--GRTVTLLNDIEGR 106
Db 866 AENLN-TLAKEHTTKGTADTALQTFKVKDGATDDEITVKGDTQNGKTWTKLKGE 924

Qy 107 PLLTV-----TATGVIOT---ROYETSSLPGRLLSVABQTPREK-----142
Db 925 NGLVATNKDGTVPFGINTQSLKAGDSTLLKGLSINKPNASNEQIQVGADGVKFAKVD 984

Qy 143 -----TSRITERLIW---AGNTEAEKDH-----NLAGQCVR 170
Db 985 KGNSTGIDGTSRITKQDQIGFTGANGSLDTTKPHLTKDKLVKGEVEIINTGINAGGKIT 1044

Qy 171 HYDTAGVTRLESILT-GTVLSQSOLLIDTQEAQNTGNETVMONLADD-----IYT 223
Db 1045 NIQSDITQNSDNDATGVRVLDKTELESKINSAAKTAQN-SLHEFSVADQGNHFTVSN 1103

Qy 224 TLSTFPA--TGALLT-----QTDAKNIQRLAYD-----VAGQLNGSWLTUKG 264
Db 1104 PYSSYDTSKTSVITFAGENGITTKVNGVVRVIGIDQTKGLTTPKLTGVNNNGKIVIDS 1163

Qy 265 QTEQVILKSL-----TYSAGQKLRBEHGNVDITEYSPEQRLIGIKTRPSDKV 317
Db 1164 KDGQNTITGLSNTLANVNDGAGHALSQGLANDT-----DKTRAAAS-----1204

Qy 318 LQDLRYEYDVPGNVI-----SIRNDAEATRFHNNQKVPENTYVDSLYQLISATGREMAN 373
Db 1205 -----IGDVLNAGFNLCNGEAVDF-----VSTYDTV-----1231

Qy 374 IGQSHQFPSPALPSDNNY--TNVTRYTYDRGNLTKIQHSSPATQNNVTNITVSNR 431
Db 1232 -----DFIDGNATTAKVYDDTSKTSKVYDVNVNDKNTIEVTSKGLGVKTTTLTKTSA 1285

Qy 432 SNRAVLSTLTEDPAQVDALFDAGGHQNTLISQNLNWNTRGELQOVTLVKRDKGAND---488
Db 1286 NGNATKFAADGADALVKA-SDIATHLNTL-----AGDIQTA-----KGASQASS 1328

Qy 489 -----DREWRYSGDG-----RRMLK-----504
Db 1329 SASVYDADGNKVIYDSTDKKYQVNDKQVDKNKEVAKDLVAQAQTPDGLTAQNNVKS 1388

Qy 505 INEQASNNNAQTRV-----TYLPNLELRL-----TONSTATTEDLQVIT-----VGEAGR 550
Db 1389 INKEQVNDANKQGINEDNAFKGLENAAKDTKTKNAAVTVGDNLNAVAQTPLTFAQDTGT 1448

Qy 551 AQVRVLHWSGKPEIDNNQLRYSVDNLIQSSQLELDSGQIISEEEYYPYGGTALWAAR 610
Db 1449 TAKKLGELTLTIKGGQTDNKLK-----DNNIG-----VWAGTD-----GFTVVKLAK 1489

Qy 611 NQTEASVYKTIYRSGKERATGLYYGYRYQYQPMIGRWLSSDPAGTIDGLNL--YRMVRNRP 669
Db 1490 DLT--NLNSVNAAGGTIRBEKGISFVDAN-----QQAANTPVL SANGLDLGGKISNIG 1541

Qy 670 VTLLDPDGL-----MPTIAERIAALKKNK-----VTDSAPSANATNVAINRPPVA 716

```

```

Db 1542 AAVDDNDVAVNPKQFNEVAKTVNNLNQSNQSGASIPFVVVTDANGKPIGTG-----KPKA 1597
Qy 717 PKPSLPK-----ASTSSQPTTHPIGAANI-----KPTTSGSSIVAPLSPVGN-----758
Db 1598 IKGADGKYHANANGVPVDKDKGKITDADKLANLAAGKPLDAGHQVVASLG--GNSDAI 1655

Qy 759 -----KST-SEISLPESAQSSSTTTSTNLOKKSFTLYRANRSPFEHQSFEPGFKAWT 812
Db 1656 TLTNIKSTLPQIDTENTGNAGAOQSLP-----SLSAAQQAASNAASVKDVLNVGFNLQT 1709

Qy 813 ---PLDTKWARQFASIFIGQKDTSNLPKETVKNISTWGAKPKLDKLSNIVKYTKDKSTVM 869
Db 1710 NNNQVD-----FKAYDTTVNFVNGT-----GA-----DITSVRSADGTMSNIT 1747

Qy 870 VSTAINTEAGQSGGAPLHKIDMDLYEFAIDGQ--KLNPL-PEGRTK-----NMVPS 918
Db 1748 VNTAL---AATDDDGNVLIK-----AKDGKFKADLMPNGSLKAGKSASDAKTPG 1796

Qy 919 LLLDTPQI---ETSSIIALNH 936
Db 1797 LSLVNPAGKSGTGDVALNN 1817

RESULT 10
US-10-467-657-1088
; Sequence 1088, Application US/10467657
; Publication No. US20050260581A1
; GENERAL INFORMATION:
; APPLICANT: CHIRON Spa
; APPLICANT: FONTANA Maria Rita
; APPLICANT: PIZZA Mariagrazia
; APPLICANT: MASIGNANI Vega
; APPLICANT: MONACI Elisabetta
; TITLE OF INVENTION: GONOCOCCAL PROTEINS AND NUCLEIC ACIDS
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/467,657
; CURRENT FILING DATE: 2003-08-11
; PRIOR APPLICATION NUMBER: GB-0103424.8
; PRIOR FILING DATE: 2001-02-12
; NUMBER OF SEQ ID NOS: 9218
; SOFTWARE: SeqWin99, version 1.04
; SEQ ID NO 1088
; LENGTH: 1468
; TYPE: PRT
; ORGANISM: Neisseria gonorrhoeae
US-10-467-657-1088

```

```

Query Match      3.1%; Score 154.5; DB 6; Length 1468;
Best Local Similarity 20.6%; Pred. No. 0.046;
Matches 213; Conservative 130; Mismatches 378; Indels 315; Gaps 55;

Qy 15 VSVYDNRGLIIRNIDFHRTTANGPDTRITRHQYDIHGHNLQSIDPRLYEAKQTNNTIKP 74
Db 111 VSVAHNGG--YNNVDFGAGSGNPD-----QHRFSY-----QIVKBNYKAGTNHPIYGG 157

Qy 75 NF-----LWQYDLTGNNPL-CTESIDAGRTVTLLNDIEGRPLLTATGVTQTRQYETSSLPG 129
Db 158 DYHMPRLHFKVTDAPVEMTSYMDGKVKYADLNKYPDR-----VRIGAGRYWRSD---207

Qy 130 RLLSVAEQTPPEKTS--RITERLIW--AGNTEAEKDH-----NLAGQCVRHVDYTAGVTRL 180
Db 208 -----EDEPNRESSVHIASAYSALVGGNTFAQNGSGGTVNLGSEKIKHSFYGLPTG 261

Qy 181 ELSLSITGTVLSQSOLLIDTQEAQNW-----TGDN-----ETVMQNNLADDIYT--223
Db 262 GSPGDSG-----SPMFIYDAQKQKWLINGVLOTGNPYIKSGNGFOLVRKDNFYDFIAGD 316

Qy 224 TLSTF-----DATGALLTQTDKAGNTQRLAYDVAGQLNGSWLTLLKQTEQVI 270
Db 317 THSVFYEPHQNGKYFFNDNNNGA--GKIDAKHKHSLPY-----RLKTRTVQLF 363

Qy 271 IKSLTYSAGQKLRBEHGNVDITEYSPEQRLIGIKTRRPSDKVLQDLRYEYDVPGN 330

```


US-11-113-424-54
; Sequence 54, Application US/11113424
; Publication No. US20050260713A1
GENERAL INFORMATION:
APPLICANT: Gangolli et al.
TITLE OF INVENTION: Polypeptides and Nucleic Acids Encoding Same
FILE REFERENCE: 21402-225
CURRENT APPLICATION NUMBER: US/11/113,424
CURRENT FILING DATE: 2005-04-21
PRIORITY APPLICATION NUMBER: 60/256,704
PRIORITY FILING DATE: 2000-12-19
PRIORITY APPLICATION NUMBER: 60/311,590
PRIORITY FILING DATE: 2001-08-10
PRIORITY APPLICATION NUMBER: 60/257,314
PRIORITY FILING DATE: 2000-12-20
PRIORITY APPLICATION NUMBER: 60/311,613
PRIORITY FILING DATE: 2001-08-10
PRIORITY APPLICATION NUMBER: 60/315,617
PRIORITY FILING DATE: 2001-08-29
PRIORITY APPLICATION NUMBER: 60/307,506
PRIORITY FILING DATE: 2001-07-24
PRIORITY APPLICATION NUMBER: 60/322,358
PRIORITY FILING DATE: 2001-09-14
PRIORITY APPLICATION NUMBER: 60/294,075
PRIORITY FILING DATE: 2001-05-29
PRIORITY APPLICATION NUMBER: 60/288,153
PRIORITY FILING DATE: 2001-05-02
NUMBER OF SEQ ID NOS: 190
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 54
LENGTH: 1045
TYPE: PRT
ORGANISM: Homo sapiens
US-11-113-424-54

Query Match 3.1%; Score 153; DB 7; Length 1045;
Best Local Similarity 19.2%; Pred. No. 0.035;
Matches 153; Conservative 99; Mismatches 242; Indels 304; Gaps 37;

Qy 15 VSVYDNR---GLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSIDPRLYEAKQTNN 70
Db 106 VNVFGRKLRVNGRNLSVDFDRTTK-----TEKIYDDH----- 138

Qy 71 TIRPNFLWQ--YDLTGNPLCTESIDAGRTVTLNDIEGRPLLVTTATGVTQTRQYETSSLP 128
Db 139 ---RKFLRLRAYDTSGHP--TLWLPSSKLMAYN-----VTYSSTGQIASIQRTGTS-- 184

Qy 129 GRLLSVAEQTPPEKTSRITERLWAGNTEAKDHNLAGQVCRHYDTAGVTRLESLSLTGT 188
Db 185 -----EKVDYDQGRIVSRVPADGKT-----WSYT 209

Qy 189 VLSQSSOLLIDTQEANWTGDNETVWQNMALDDIYTLTSTFDTATGALLTQTDAGKNIQRLA 248
Db 210 YLEKSMVLLHSOR-----QVIFEYDMWDRLS-----AITMPSVARHTMQ--- 249

Qy 249 YDVAGQLNGSWLTKGQTEQVVIKSLTYSAAQKLRHEHGNVDVITEYSYEP-----E 300
Db 250 -----TIRSIGYRNINPPESNAS--IITDYNEEGLLQTAFLG 287

Qy 301 TORLIGIKTRRPD--TKVLQD---LRYEYDPVGNVISIRN-----D 337
Db 288 TSSRVLPKYRQTRLSILYDSTRVSTFYDETAGVLTQVNLQSDGFICTIRYRQIGPLID 347

Qy 338 AEATRFHNNQKVPENTYVDSYQLISATG-----REMANIGQQSHQP----- 382
Db 348 RQIFRPSDEGMVARNPDSYSDNSFRVTSMQGVINETPLDILYQFDDISGKVEQGFQFV 407

Qy 383 -----SPALPSDNNVTN-----YTRYVYDRCGNLTKIQHSSPA 417
Db 408 IYDINDQIISAVMTYTKFPAHGRKIEIQEIFRSLMTWITIQDYNMGRVTKREIKIGP 467
Db 418 TQN-----NY--TTNITVSNKSNRAVLSTLTD----- 443

RESULT 13
US-10-821-234-1097
; Sequence 1097, Application US/10821234
; Publication No. US20050255114A1
GENERAL INFORMATION:
APPLICANT: Labat, Ivan
APPLICANT: Stache-Crain, Birgit
APPLICANT: Andarmani, Susan
APPLICANT: Tang, Y. Tom
TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
FILE REFERENCE: 821A
CURRENT APPLICATION NUMBER: US/10/821,234
CURRENT FILING DATE: 2004-04-07
PRIORITY APPLICATION NUMBER: US 60/462,047
PRIORITY FILING DATE: 2003-04-07
NUMBER OF SEQ ID NOS: 1704
SOFTWARE: pt_seq_genes Version 1.0
SEQ ID NO 1097
LENGTH: 1094
TYPE: PRT
ORGANISM: Homo sapiens
US-10-821-234-1097

Query Match 3.1%; Score 153; DB 6; Length 1094;
Best Local Similarity 19.2%; Pred. No. 0.037;
Matches 153; Conservative 99; Mismatches 242; Indels 304; Gaps 37;

Qy 15 VSVYDNR---GLIIRNIDPHRTTANGDPDTRITRHOYDIHGLNQSIDPRLYEAKQTNN 70
Db 155 VNVFGRKLRVNGRNLSVDFDRTTK-----TEKIYDDH----- 187

Qy 71 TIRPNFLWQ--YDLTGNPLCTESIDAGRTVTLNDIEGRPLLVTTATGVTQTRQYETSSLP 128
Db 188 ---RKFLRLRAYDTSGHP--TLWLPSSKLMAYN-----VTYSSTGQIASIQRTGTS-- 233

Qy 129 GRLLSVAEQTPPEKTSRITERLWAGNTEAKDHNLAGQVCRHYDTAGVTRLESLSLTGT 188
Db 234 -----EKVDYDQGRIVSRVPADGKT-----WSYT 258

Qy 189 VLSQSSOLLIDTQEANWTGDNETVWQNMALDDIYTLTSTFDTATGALLTQTDAGKNIQRLA 248
Db 259 YLEKSMVLLHSOR-----QVIFEYDMWDRLS-----AITMPSVARHTMQ--- 298

Qy 249 YDVAGQLNGSWLTKGQTEQVVIKSLTYSAAQKLRHEHGNVDVITEYSYEP-----E 300
Db 299 -----TIRSIGYRNINPPESNAS--IITDYNEEGLLQTAFLG 336

Qy 301 TORLIGIKTRRPD--TKVLQD---LRYEYDPVGNVISIRN-----D 337
Db 337 TSSRVLPKYRQTRLSILYDSTRVSTFYDETAGVLTQVNLQSDGFICTIRYRQIGPLID 396


```
; LENGTH: 2715  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-11-096-051-2
```

Query Match 3.1%; Score 153; DB 7; Length 2715;
Best Local Similarity 19.2%; Pred. No. 0.14;
Matches 153; Conservative 99; Mismatches 242; Indels 304; Gaps 37;

Qy	15	VSVYDNR----	GLIIRNIDFHRTTANGDPDRITRHOYDIHGHLNOSIDPRLYEAKQTNN	70
Dd	1776	VNVFGRKLRVNGRLLSVDFRITK-----TEKIYDDH-----	1808	
Qy	71	TIKPNFLWQ--YDLTGNPCLTESIDAORTVTLNDIEGRPLLTWTATGVIOTROQETSSLP	128	
Dd	1809	--RKFTLLRAYDTSGHP--TLMPFSKLMAVN-----VTYSSTQIASIQGTTS--	1854	
Qy	129	GRLLSVAEQPEEKTSITIERLIWAGNTAEAKOHNLAGQCVRHYDVAGTVRLSLSLTGT	188	
Dd	1855	-----EKVYDGQRIVSRVFADGKT-----WSYT	1879	
Qy	189	VLSOSSOLLIDTOEANWGDNETWQNMLADDIVTTLSFTFDGALLTOTDAKNIGORLA	248	
Dd	1880	YLEKSMVLILHSQR-----QIFBYDMWRLS-----AIMPVSARHTWQ---	1919	
Qy	249	YDVAGQLNGSWLTKGQTEOVIIKSULTSYSAAGOKREEHGNDIVITEYSYP-	E 300	
Dd	1920	-----TIISICYERNINPPESNAS-IITDYNEBGLLLQTAFLG	1957	
Qy	301	TORLIGHKTRRPD-TKVLOD---LRVEYPDGVNGVISIRN-----D	337	
Dd	1958	TSRVLPKYRRQTRLSEILYSTRVSTYDETAGLVKTNVLDQSDGFICTIRYRIQIGLID	2017	
Qy	338	ABATRFEHWNKVPENTTYDSLYQLLSATG-----REMANIGOOSHQP----	382	
Dd	2018	RQIFRSEDGWNARFYSDNSFRVTSMOGINETPLPDILYQFDDISKGVEQFKFGV	2077	
Qy	383	----SPALPSDNNITYN-----YTRTTYDRGGNLTKIOHSPA	417	
Dd	2078	IYYDINGOIISTAVNTYTHFDAHCRIKEIOYEIFRSIMYWITTIQYDNMGRVTKREIKGP	2137	
Qy	418	TQN-----NY--TMTITVSNRSRAVSLTFED-----	443	
Dd	2138	FANTTKVAYEYVDQGLOTQTVYLNEKIMRWYNVDLNGNLHLNFNSARLTPLYDLRDRI	2197	
Qy	444	--PAQWDALFDAGC---HQWTLISGOQLNWNTRGELQQVTLVKDKGANDDRWW---	YRY 495	
Dd	2198	TRLGDVOYRLDEDFGLRQGTGEI-----FEYSSKGLLTRV-----YSKGS-	2245	
Qy	496	SGGRRMLKINEQAASNAQTO-----RVLYLPNLLELRLTONSTATTEDELQVITV	545	
Dd	2246	DGLGRRV-----SSKTSLGQHQLFPFYADLTPTPLTRIITHVIN---HSSSEITSLYYDLQ---	2294	
Qy	546	GEAGRAQVRVLHWESGKPEDINDNOLRYSYDNLIGSSOLEJDESGOIIISEEEYYPYGGA	605	
Dd	2295	-----GHLPAMEISSG-----DEFYIASDN-TGTPLAYPSNGLMKLQIQTATAGEIY	2341	
Qy	606	LWAARNOTEASYKTI--RYSCKERDATGLYYYGYRYYQPWIGRW-----LSSDPA	653	
Dd	2342	F-----DSNIDFQLVIGFHGGLYDPLKLIHFGERDYDILAGRWTTPDIBIWKRICKOPA	2396	
Qy	654	GTIDGLNLRYMRVRNPNVT	671	
Dd	2397	-----PFNLNMFRRNNPNPAS	2410	

Search completed: February 16, 2006, 21:55:22
Job time : 12.1786 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:22:44 ; Search time 95.1891 Seconds
(without alignments)
4431.220 Million cell updates/sec

Title: US-10-754-115-47

Perfect score: 5005

Sequence: 1 MKNDPKLYQKTPVSYVDN.....DAEISFLTTIPLKNVXPKHR 960

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*

- 1: Geneseqp1980s.*
- 2: Geneseqp1990s.*
- 3: Geneseqp2000s.*
- 4: Geneseqp2001s.*
- 5: Geneseqp2002s.*
- 6: Geneseqp2003as.*
- 7: Geneseqp2003bs.*
- 8: Geneseqp2004s.*
- 9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5005	100.0	960	8	ADP18624 Photorhab
2	5005	100.0	960	8	ADR21527 Photorhab
3	4545	90.8	954	2	AAY33728 Photorhab
4	3300.5	65.9	971	6	ABM70223 Photorhab
5	2879	57.5	938	8	ADP18628 Photorhab
6	2879	57.5	938	8	ADR21579 Photorhab
7	2850.5	57.0	949	8	ADP18626 Photorhab
8	2850.5	57.0	949	8	ADR21544 Photorhab
9	2833.5	56.6	939	6	ABM70226 Photorhab
10	2821.5	56.4	960	6	ABM70214 Photorhab
11	2768.5	55.3	966	6	ABM69154 Photorhab
12	2744	54.8	1044	6	ABM69055 Photorhab
13	2740	54.7	1043	2	AAW18306 Photorhab
14	2740	54.7	1043	2	AAW56569 Toxin Tcc
15	2740	54.7	1043	6	ABM70231 Photorhab
16	2539	50.7	916	6	ABM70231 Photorhab
17	2537.5	50.7	915	5	ABG32853 F. lumine
18	2535.5	50.7	915	8	ADR21538 Photorhab
19	2383	47.6	937	6	ABM67982 Photorhab
20	2305	46.1	971	3	AAY95687 Cosmid cH
21	2304.5	46.0	962	8	ADR21531 Xenorhabd
22	2304.5	46.0	962	9	ABE47814 Native Xp
23	2226	44.5	1016	8	ADR20367 Recombina
24	2226	44.5	1016	8	ADR21496 Xenorhabd

25	2184	43.6	973	4	AAY97696	Aay97696 SepC prot
26	2184	43.6	973	8	ADR21541	Adr21541 Serratia
27	1581.5	31.6	930	8	ADN61349	Adn61349 Paenibaci
28	1581.5	31.6	930	8	ADR21522	Adr21522 Paenibaci
29	1581.5	31.6	930	8	ADN61355	Adn61355 Paenibaci
30	1581.5	31.6	953	8	ADR21523	Adr21523 Paenibaci
31	1268.5	25.3	940	7	ADG73071	Adg73071 Pseudomon
32	1268.5	25.3	940	7	ADL12126	Adl12126 Pseudomon
33	651	13.0	286	8	ADN61377	Adn61377 Paenibaci
34	332.5	6.6	932	6	ABU16693	Abu16693 Protein e
35	329.5	6.6	1572	6	ABU41491	Abu41491 Protein e
36	329.5	6.6	1627	6	ADA35317	Ada35317 Acinetoba
37	322.5	6.4	998	7	ABO79678	AbO79678 Pseudomon
38	321	6.4	1530	6	ABU40251	Abu40251 Protein e
39	318	6.4	1385	6	ABU40318	Abu40318 Protein e
40	317	6.3	1565	6	ABU16634	Abu16634 Protein e
41	315	6.3	1317	8	ADS22210	Ads22210 Bacterial
42	308	6.2	2334	5	ABG31849	Abg31849 Human kin
43	307.5	6.1	843	6	ABU48290	Abu48290 Protein e
44	303.5	6.1	1439	7	ADF07275	Adf07275 Bacterial
45	302.5	6.0	1253	6	ABU40990	Abu40990 Protein e

ALIGNMENTS

RESULT 1

ADP18624

ID ADP18624 standard; protein; 960 AA.

XX AC ADP18624;

XX DT 12-AUG-2004 (first entry)

XX DE Photorhabdus luminescens W-14 tccc3 protein SEQ ID NO:12.

XX KW tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;

XX KW orally active insect toxin; insect toxin; tccc3.

XX OS Photorhabdus luminescens.

XX PN WO2004044217-A2.

XX PD 27-MAY-2004.

XX PF 12-NOV-2003; 2003WO-IB005553.

XX PR 12-NOV-2002; 2002US-0425672P.

XX (UYBA-) UNIV BATH.

XX Ffrench-Constant RH, Waterfield NR;

XX WPI; 2004-411735/38.

XX N-PsDB; ADP18623.

New isolated Photorhabdus luminescens nucleic acids, useful for expressing orally active insect toxin or for generating transgenic plants with enhanced resistance to insects.

Claim 1; SEQ ID NO 12; 118pp; English.

The present invention describes DNA sequences from the tcd genomic region of Photorhabdus luminescens W-14. Also described: (1) a transgenic monocot or dicot cell having a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant with a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a progeny of seed; (5) a method of producing Toxin A of Photorhabdus luminescens W-14 in a heterologous host; and (6) a method of producing an orally active insect toxin. The nucleotide sequences are useful for heterologous expression of orally active insect toxin. They can also be used for generating transgenic plants with enhanced resistance to

CC insects. The present sequence represents Photorhabdus luminescens W-14
CC tccc3, which is used in the exemplification of the present invention.
XX
SQ Sequence 960 AA;

Query Match 100.0%; Score 5005; DB 8; Length 960;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60
Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60

Qy 61 RLYEAKQNTNNTIKNFWLQYDLTGKPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120
Db 61 RLYEAKQNTNNTIKNFWLQYDLTGKPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120

Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180

Qy 181 ELSLITGTVLSSQSSQLLIDTQGANWGTGNETWQNLADDIYTTLTSTPDATGALLTQDGA 240
Db 181 ELSLITGTVLSSQSSQLLIDTQGANWGTGNETWQNLADDIYTTLTSTPDATGALLTQDGA 240

Qy 241 KGNITQRLAYDVAGQLNGSWLTLKGTEQVVIKSLTYSAGOKLREEHGNDVITEVSEPE 300
Db 241 KGNITQRLAYDVAGQLNGSWLTLKGTEQVVIKSLTYSAGOKLREEHGNDVITEVSEPE 300

Qy 301 TORLIGIKTRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWHNQKVPENTYTYDSL 360
Db 301 TORLIGIKTRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWHNQKVPENTYTYDSL 360

Qy 361 YQLISATGREMANIQQSHQSPALPSDNNYTYTFTYTYDRGNNLTKIQHSPATQN 420
Db 361 YQLISATGREMANIQQSHQSPALPSDNNYTYTFTYTYDRGNNLTKIQHSPATQN 420

Qy 421 NYTNTITVSNRNLAVLTSLTDPADQVLDALPDAGHONTLSGQNLNNTREGELQOVTLV 480
Db 421 NYTNTITVSNRNLAVLTSLTDPADQVLDALPDAGHONTLSGQNLNNTREGELQOVTLV 480

Qy 481 KRDKGANDREWIRYSGDGRMLKINEQQAANNAQTRVTLPLNLELRLTQNSTATTEDL 540
Db 481 KRDKGANDREWIRYSGDGRMLKINEQQAANNAQTRVTLPLNLELRLTQNSTATTEDL 540

Qy 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEEEYYP 600
Db 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISSEEEYYP 600

Qy 601 YGGTALWAARQTEASYKTIYSGKERDATGLYGYRYOPWIGRWLSSDPAGTIDGLN 660
Db 601 YGGTALWAARQTEASYKTIYSGKERDATGLYGYRYOPWIGRWLSSDPAGTIDGLN 660

Qy 661 LYRWVRNPNVTLDPDGLMPTIAERIAALKKKNKVTDSAPSPANATVAINIRPPVAPKPS 720
Db 661 LYRWVRNPNVTLDPDGLMPTIAERIAALKKKNKVTDSAPSPANATVAINIRPPVAPKPS 720

Qy 721 LPKASTSSQPTHTPIGAANIPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSTTST 780
Db 721 LPKASTSSQPTHTPIGAANIPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSSTTST 780

Qy 781 NLQKKSFTLYRADNRSPEEMOSKPEPGKATPLDTKWAROPASIFIGQKDTSNLPKETV 840
Db 781 NLQKKSFTLYRADNRSPEEMOSKPEPGKATPLDTKWAROPASIFIGQKDTSNLPKETV 840

Qy 841 KNIITWGAAPKLKOLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
Db 841 KNIITWGAAPKLKOLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900

Qy 901 GQKLNPLPEGTKNMPVSLDLDTPQIETSSIIALNHGVPNDABISFLTPIPLKNVKKPHR 960
Db 901 GQKLNPLPEGTKNMPVSLDLDTPQIETSSIIALNHGVPNDABISFLTPIPLKNVKKPHR 960

RESULT 2
ADR21527
ID ADR21527 standard; protein; 960 AA.
XX
AC ADR21527;
XX
DT 04-NOV-2004 (first entry)
XX
DE Photorhabdus Tccc3 toxin SEQ ID NO:47.
XX
KW toxin; insect; insecticidal; transgenic; pest control.
XX
OS Photorhabdus luminescens.
XX
FN WO2004067727-A2.
XX
PD 12-AUG-2004.
XX
PF 07-JAN-2004; 2004WO-US000394.
XX
PR 21-JAN-2003; 2003US-0441723P.
XX
FA (DOWC) DOW AGROSCIENCES LLC.
XX
PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
XX
DR WPI; 2004-580999/56.
DR N-PSDB; ADR21526.
XX
PT Controlling or inhibiting an insect, useful for pest control, comprises
PT contacting the insect with effective amounts of a Protein A, a Protein B,
PT and a Protein C.
XX
PS Claim 1; SEQ ID NO 47; 368pp; English.
XX
CC The invention relates to a novel method for controlling or inhibiting an
CC insect comprising contacting the insect with effective amounts of a
CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
CC is encoded by a naturally occurring gene or has an amino acid sequence
CC that differs from the product encoded by a naturally occurring gene only
CC by truncation or by conservative amino acid changes. Protein A is a 230-
CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic
CC species, has stand alone insecticidal activity, and has an amino acid
CC sequence at least 40% identical to a sequence selected from XptAlwi,
CC XptA2w1, TcdA, TcdA2, TcdA4, and TcdA. Protein B is a 130-180 kDa toxin
CC complex potentiator having an amino acid sequence at least 40% identical
CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1w1, XptB1xb,
CC PptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
CC having an amino acid sequence at least 35% identical to a sequence
CC selected from Tccc1, Tccc2, Tccc3, Tccc4, Tccc5, XptB1w1, XptC1xb, PptC1
CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic
CC plant or plant cell that produces a Protein A, a Protein B, and a Protein
CC C. The method is useful for pest control. The present sequence represents
CC Photorhabdus luminescens Tccc3 toxin.
XX
SQ Sequence 960 AA;

Query Match 100.0%; Score 5005; DB 8; Length 960;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 960; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60
Db 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGLNQSIDP 60

Qy 61 RLYEAKQNTNNTIKNFWLQYDLTGKPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120
Db 61 RLYEAKQNTNNTIKNFWLQYDLTGKPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120

Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180
Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAEKDNHLAGQCVRHDTAGVTRL 180

Db 121 QYETSSLPGRLLSVAEQTPPEKTSRTERLIWAGNTEAEKHNLGAGCVRHYDTAGVTRL 180
 QY 181 ESSLTGTVLSSQSSQLLIDTOEANTGNQNETVWQNLADDIYTTLSFTDAGTALLTQIDA 240
 Db 181 ESSLTGTVLSSQSSQLLIDTOEANTGNQNETVWQNLADDIYTTLSFTDAGTALLTQIDA 240
 QY 241 KGNIRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHGNVDITEYSYBPE 300
 Db 241 KGNIRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHGNVDITEYSYBPE 300
 QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWQNKVMPENTTYDLSL 360
 Db 301 TORLIGIKTRRPSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWQNKVMPENTTYDLSL 360
 QY 361 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 420
 Db 361 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 420
 QY 421 NYTTNITVSNRRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTRGELQOQVTLV 480
 Db 421 NYTTNITVSNRRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTRGELQOQVTLV 480
 QY 481 KDKGANDREWRYSGDGRMLKINEQOASNAQTORVYLPNLELRLTONSTATTEDL 540
 Db 481 KDKGANDREWRYSGDGRMLKINEQOASNAQTORVYLPNLELRLTONSTATTEDL 540
 QY 541 QVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLDSEGQIISBEEYYP 600
 Db 541 QVITVGEAGRAQVRVLHWSGKPEDIDNNQLRYSYDNLIGSSQLDSEGQIISBEEYYP 600
 QY 601 YGGTALWAARQTEASYKTIYSGKRDATGLYYGYRYQPMIGRWLSSDPAGTIDGLN 660
 Db 601 YGGTALWAARQTEASYKTIYSGKRDATGLYYGYRYQPMIGRWLSSDPAGTIDGLN 660
 QY 661 LYRWVRNPVTLDPDGLMPTTAERIALKKNKVTDSPSPANATNVAIRPPVAPKPS 720
 Db 661 LYRWVRNPVTLDPDGLMPTTAERIALKKNKVTDSPSPANATNVAIRPPVAPKPS 720
 QY 721 LPKASTSQPTTHPIGANAIPKPTSSGSSIVAPLSPVGNKSTSEISLPESAQSSSSSTTST 780
 Db 721 LPKASTSQPTTHPIGANAIPKPTSSGSSIVAPLSPVGNKSTSEISLPESAQSSSSSTTST 780
 QY 781 NLQKSFLLYRADNRSFEMQSKFPEGFKAWTPLDTKWARQFASIFIGQKOTSNLPKETV 840
 Db 781 NLQKSFLLYRADNRSFEMQSKFPEGFKAWTPLDTKWARQFASIFIGQKOTSNLPKETV 840
 QY 841 KNIISTWGAAPKLKOLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
 Db 841 KNIISTWGAAPKLKOLSNYIKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFAID 900
 QY 901 GQKLNPLPEGRTKNMPVSLLDTPQIETSSIIALNHGPNDAEISFLTTIPLKNVKPKHR 960
 Db 901 GQKLNPLPEGRTKNMPVSLLDTPQIETSSIIALNHGPNDAEISFLTTIPLKNVKPKHR 960

RESULT 3
 ID AAY33728 standard; protein; 954 AA.
 XX AAY33728;
 AC AAY33728;
 XX 09-NOV-1999 (first entry)
 DT Photorhabdus luminescens 954 amino acid insecticidal toxin.
 DE Symbiotic bacterium; nematode; insect; larva; toxin; insecticide.
 KW Photorhabdus luminescens.
 OS Photorhabdus luminescens.
 XX WO9942589-A2.
 PN 26-AUG-1999.
 XX

PF 18-FEB-1999; 99WO-EP001015.
 XX 20-FEB-1998; 98US-00027080.
 PR 20-JAN-1999; 99US-0116439P.
 XX (NOVS) NOVARTIS AG
 PA (NOVS) NOVARTIS-ERFINDUNGEN VERW GES MBH.
 XX Kramer VC, Morgan MK, Anderson AR, Hart HP, Warren GW, Dunn MM;
 PI Chen JS;
 XX WPI; 1999-527479/44.
 DR N-PSDB; AAZ06831.
 XX New nucleic acid from Photorhabdus luminescens encoding insecticidal
 PT toxins, used for making resistant transgenic plants.
 XX Claim 26; Page 130-133; 148pp; English.
 PS This sequence represents a 954 amino acid insecticidal toxin from
 CC Photorhabdus luminescens. It is one of three insecticidal toxins
 CC (AAY33728-Y33730) encoded by open reading frames (orfs) in a 38kb
 CC fragment of P. luminescens DNA (AAZ06831). This sequence is encoded by
 CC orf5. P. luminescens is a member of the Enterobacteriaceae family and is
 CC a symbiotic bacterium of nematodes of the genus Heterorhabditis. The
 CC nematodes colonise insect larvae, kill them, and their offspring feed on
 CC the dead larvae. However, the insecticidal agents are produced by P.
 CC luminescens rather than the nematodes. The toxins have activity against
 CC Lepidopteran insects such as Cabbage Looper (Trichoplusia ni), European
 CC Corn Borer (Ostrinia nubilalis) and Fall Armyworm (Spodoptera frugiperda)
 CC and also against Coleopteran insects (e.g., Colorado Potato Beetle,
 CC Leptinotarsa decimlineata). In addition the toxins are active against
 CC strains resistant to known insecticides. The DNA sequence can be used to
 CC generate transgenic plants of various species that are resistant to
 CC economically important insect pests and also for recombinant production
 CC of the toxins for use as insecticides
 XX Sequence 954 AA;

Query Match 90.8%; Score 4545; DB 2; Length 954;
 Best Local Similarity 90.4%; Pred. No. 3.6e-307;
 Matches 868; Conservative 40; Mismatches 46; Indels 6; Gaps 2;
 QY 1 MKNIDPKLYQKTPVSVVYDNRGLIIRNIDFHRITANGDPDTRITRHQYDIHGLHQSIDP 60
 Db 1 MKNIDPKLYQKTPVSVVYDNRGLIIRNIDFHRITANGDPDTRITRHQYDIHGLHQSIDP 60
 QY 61 RLYEAKQNTNNTIKNPLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
 Db 61 RLYEAKQNTNNTIKNPLWQYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 115
 QY 121 QYETSSLPGRLLSVAEQTPPEKTSRTERLIWAGNTEAEKHNLGAGCVRHYDTAGVTRL 180
 Db 116 QYETSSLPGRLLSVAEQTPPEKTSRTERLIWAGNTEAEKHNLGAGCVRHYDTAGVTRL 174
 QY 181 ESSLTGTVLSSQSSQLLIDTOEANTGNQNETVWQNLADDIYTTLSFTDAGTALLTQIDA 240
 Db 175 ESSLTGTVLSSQSSQLLIDTOEANTGNQNETVWQNLADDIYTTLSFTDAGTALLTQIDA 234
 QY 241 KGNIRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHGNVDITEYSYBPE 300
 Db 235 KGNIRLAYDVAGQLNGSWLTKGTEQVVIKSLTYSAGOKLREHGNVDITEYSYBPE 294
 QY 301 TORLIGIKTRRPSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWQNKVMPENTTYDLSL 360
 Db 295 TORLIGIKTRRPSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWQNKVMPENTTYDLSL 354
 QY 361 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 420
 Db 355 YQLISATGREMANIQSQSHQFPSPALPSDNNNTYNTYTYTYDRCGNLTQKHSSPATON 414
 QY 421 NYTTNITVSNRRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTRGELQOQVTLV 480

Db 415 NYTNTITVSNRRAVLSTLTEDPAQVDALFDAGGHONTLISGQNLNWNTRGELQHVTLV 474

Qy 481 KRDGANDREWYSGRRMLKINEQOASNAQOTQVTVLPNLELRLTONSTATTEDL 540

Db 475 KRDGANDREWYSGRRMLKINEQOASNAQOTQVTVLPNLELRLTONSTATTEDL 534

Qy 541 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIEEYYP 600

Db 535 QVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIEEYYP 594

Qy 601 YGGTALMAARNQTEASYKTIYSGKERDATGLYYGYRYYPQWIGRLSSDPAGTIDGLN 660

Db 595 YGGTALMAARNQTEASYKTIYSGKERDATGLYYGYRYYPQWIGRLSSDPAGTIDGLN 654

Qy 661 LYRVVRNPNVTLDPDGLMPTIAERIALKKNKYVTDSPANATNVAINIRPPVAKPS 720

Db 655 LYRVVRNPNVTLDPDGLMPTIAERIALKKNKYVTDSPANATNVAINIRPPVAKPS 714

Qy 721 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSTTST 780

Db 715 LPKASTSQPTTHPIGAANIKPTTSGSSIVAPLSPVGNKSTSEISLPSAQSSSTTST 774

Qy 781 NLOKGSFTLYRADNRSEEMQSKPEGPKAWTPLDTKMARQFASIFIGQKDTSNLPKXTV 840

Db 775 NLOKGSFTLYRADNRSEEMQSKPEGPKAWTPLDTKMARQFASIFIGQKDTSNLPKXTV 834

Qy 841 KNIISTWGAKKPLKDLNYSIKYTKDSTVWVSTAINTEAGSGGAPLHKIDMDLYEPAID 900

Db 835 KNIINTWGTGPKLNDLSTYIKYTKDSTVWVSTAINTEAGSGGAPLHKIDMDLYEPAID 894

Qy 901 QOKNPLPEGTKNWPSLLDTPQIETSSIIALNHGPNDAEISFLLTTPKKNVKKPKR 960

Db 895 QOKNPLPRGSKORVPSLLDTPQIETSSIIALNHGPNDAEISFLLTTPKKNVKKPKR 954

RESULT 4

ABM70223

ID ABM70223 standard; protein; 971 AA.

AC ABM70223;

XX 20-NOV-2003 (first entry)

DT 20-NOV-2003

XX Photorhabdus luminescens protein sequence #3320.

DE

XX Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;

KW detection; food; gene expression; plant; animal; microorganism; toxin;

KW antibiotic; biopesticide; virulence factor; disease model; plague;

XX whooping cough.

XX Photorhabdus luminescens.

OS

XX WO200294867-A2.

PN

XX 28-NOV-2002.

PD

XX 07-FEB-2002; 2002WO-IB003040.

PF

XX 07-FEB-2001; 2001FR-00001659.

PR

XX (INSP) INST PASTEUR.

PA (CNRS) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

PI Buchrieser C;

XX WPI; 2003-148459/14.

DR

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,

PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX Claim 2; SEQ ID NO 3320; 1205pp; French.

PS

XX

CC The invention relates to the isolation of genes and their encoded

CC proteins from Photorhabdus luminescens. The isolated sequences are

CC sources of probes and primers for detecting the genome of P. luminescens

CC and related species; to study polymorphisms; for gene analysis and for

CC detection/amplification of the genes. Antibodies (Ab) raised against the

CC polypeptides encoded by the genes are used for detection/identification

CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that

CC carry a gene-containing vector are used to select compounds that

CC modulate, regulate, induce or inhibit expression of the genes in plants,

CC animals or microorganisms other than P. luminescens and are able to alter

CC response or sensitivity to toxins and antibiotics produced by P.

CC luminescens. Cells transformed to express the genes are useful for

CC recombinant production of the proteins, particularly toxins and

CC antibacterials useful as insecticides, bactericides and fungicides. The

CC genes, proteins, vectors containing the genes and Ab are also useful

CC therapeutically to treat microbial infection by bacteria or fungi that

CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as

CC biopesticides. Other uses of the genes and the proteins are as virulence

CC factors and for identifying targets of human diseases for which P.

CC luminescens is a model (particularly plague and whooping cough). This

CC sequence represents one of the isolated P. luminescens proteins

XX

SQ Sequence 971 AA;

Query Match 65.9%; Score 3300.5; DB 6; Length 971;

Best Local Similarity 67.6%; Pred. No. 1.7e-220;

Matches 679; Conservative 81; Mismatches 151; Indels 93; Gaps 17;

Qy 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRITANGDPTTRITRHOYDIGHLNSIDP 60

Db 1 MKNIDPKLYOKTPTVSVYDNRGLIIRNIDFHRITANGDPTTRITRHOYDIGHLNSIDP 60

Qy 61 RLYEAKTNTNLIKFNFLWOYDLTGNIPLCTESIDAGRTVTLNDIEGRPLLTVTATGVQTR 120

Db 61 RLYEAKTNTNLIKFNFLWOYDLTGNIPLCTESIDAGRTVTLNDIEGRPLLTVTATGVQTR 120

Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGTEAKDHNLAGOCVRHYDTAGVTRL 180

Db 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGTEAKDHNLAGOCVRHYDTAGVTRL 180

Qy 181 ESSLTGTVLSSQSLLDITQEBANWTGDNETWQWMLADDIYTLTSTFDATGALLTQDA 240

Db 181 ESSLTGTVLSSQSLLDITQEBANWTGDNETWQWMLADDIYTLTSTFDATGALLTQDA 240

Qy 241 KGNIRLAYVAGOLNGSLTLKGTOBOVIKSLTYSAAQKREEHGNDVITEYSYEPE 300

Db 241 KGNIRLAYVAGOLNGSLTLKGTOBOVIKSLTYSAAQKREEHGNDVITEYSYEPE 300

Qy 301 TORLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFHKNQKVPENTY 355

Db 301 TORLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFHKNQKVPENTY 360

Qy 356 TYDSLQYLIISATGREMANIGQOSHQFPSPALPSDNNYTYNTRYTYTDRGGLTKIQHSS 415

Db 361 IYDSLQYLIISATGREMANIGQOSHQFPSPALPSDNNYTYNTRYTYTDRGGLTKIQHSS 420

Qy 416 PATQNNYTYNTRYTYTDRGGLTKIQHSS 475

Db 421 PATQNNYTYNTRYTYTDRGGLTKIQHSS 480

Qy 476 QVTLVKRDKGANDREWYSGDGRRLKINEQOASNAQOTQVTVLPNLELRLTONSTATTEDL 535

Db 481 QVTLVKRDKGANDREWYSGDGRRLKINEQOASNAQOTQVTVLPNLELRLTONSTATTEDL 540

Qy 536 TTEDLOVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIEEYYP 595

Db 541 TTEDLOVITVGEAGRAQVRLVHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIEEYYP 600

Qy 596 BEYYPYGGTALMAARNQTEASYKTIYSGKERDATGLYYGYRYYPQWIGRLSSDPAGT 655

Db 601 BEYYPYGGTALMAARNQTEASYKTIYSGKERDATGLYYGYRYYPQWIGRLSSDPAGT 660

Qy 656 IDGLNLYRMVRNPNVTLDPDGLMPTIAE-----RIAAKKNKVTDS 697

[illegible]

RESULT 5

ADP18628	
ID	ADP18628 standard; protein; 938 AA.
XX	
AC	ADP18628;
XX	
DT	12-AUG-2004 (first entry)
XX	
DE	Photorhabdus luminescens W-14 tccC5 protein SEQ ID NO:16.
XX	
KW	tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;
KW	orally active insect toxin; insect toxin; tccC5.
XX	
OS	Photorhabdus luminescens.
XX	
FN	WO200404217-A2.
XX	
PD	27-MAY-2004.
XX	
Pf	12-NOV-2003; 2003WO-IB005553.
XX	
PR	12-NOV-2002; 2002US-0425672P.
XX	
PA	(UYBA-) UNIV BATH.
XX	
Pi	Ffrench-Constant RH, Waterfield NR;
XX	
DR	WPI; 2004-411735/38.
DR	N-PsDB; ADP18627.
XX	
PT	New isolated Photorhabdus luminescens nucleic acids, useful for
PT	expressing orally active insect toxin or for generating transgenic plants
PT	with enhanced resistance to insects.
XX	
PS	Claim 1; SEQ ID NO 16; 118pp; English.

The present invention describes DNA sequences from the tcd genomic region of *Photobacterium luminescens* W-14. Also described: (1) a transgenic monocol or dicot cell having a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant with a genome comprising a nucleic acid sequence that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a progeny of seed; (5) a method of producing Toxin A of *Photobacterium luminescens* W-14 in a heterologous host; and (6) a method of producing an orally active insect toxin. The nucleotide sequences are useful for heterologous expression of orally active insect toxin. They can also be used for generating transgenic plants with enhanced resistance to insects. The present sequence represents *Photobacterium luminescens* W-14 tcd5c3, which is used in the exemplification of the present invention.

XX	SQ	Sequence	938 AA;
		Query Match	57.5%; Score 2879; DB 8; Length 938;
		Beat Local Similarity	60.9%; Pred. No. 3.7e-191;
		Matches	591; Conservative 90; Mismatches 186; Indels 104; Gaps 16
Qy	1	MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHTRTANGDPDTRITHQYDIHGHLNQSIDP	60
Db	1	MENIDPKLYHHTPTVSVHDNRGLAIRNISFRTTAEANTDTRITHQYNAGLYNQSIDP	60
Qy	61	PLYEAKQFNNTIKNFLWOYDLTGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIOIR	120
Db	61	RLYDAKQFNNAVQNFIRHNLITGNILKTESVDAGRTITLNDIEGRPVLTINAAVRQNH	120
Qy	121	QYETSSPLGRLLSVAEQ--TPBEKTSRIERLIWAGNTAEAKDHNLAGQCVRHVDYTAGVTR	179
Db	121	RYEDNTLPGRLLASEQQAEBKT--TERLIWAGNTPOEKDHNLAGQCVRHVDYTAGLQ	177
Qy	180	LESLSLTGTVLSQSQLLIDTOEANWTVGNQNTVWQNMLADDIYTLTSTPDATGALLTQTD	239
Db	178	LNSLALTGAVLSSQSQQLLTDNQADWTGEDQSLWQOKLSSDVIYTSQNTSDATGALLTQTD	237
Qy	240	AKGNIQRLAYDVGOLNGSWLTLAGQTVQVTKSLTYSNAGOKLKEEHCNDVITEYSYEP	299
Db	238	AKGNIQRLAYDVGOLKGSWLTLAGQAEQVTKSLTYSNAGOKLKEEHCNGIVITEYSYEP	297
Qy	300	ETQRLIGIKTRRPSDKVLQDLRYEYDPVGNVISIRNDAEATRFWNQKVPENTVITYDS	359
Db	298	ETQRLIGIKTRRPSDAKVLQDLRYQYDPVGNVISIRNDAEATRFWNQKVAPEVITYDS	357
Qy	360	LYQLISATGREMANIQGQSHQPPSPALPSDNNYTNTRYTYDREGNLTKIQHSPATQ	419
Db	358	LYQLISATGREMANIQGQSNQLPSPALPSDNNYTNTRYTYDREGNLTKIQHSPAAQ	417
Qy	420	NNYTNITVSRNRAVLSTLTEDPAQVDALPDAGGHQNTLLSGQNLNWTNRTGELQOVTL	479
Db	418	NNYTTDITVSRNRAVLSTLTADPTQVDALPDAGGHQNTLLSGQVLTWTPREGELQ---	474
Qy	480	VKRDKGANDD--REWYRYSGGRRMLKNEOQASNNQAOTQRTVYLPNLELRITONSTATT	537
Db	475	-----ANNSAGNEWRYDSDNGIRQLKNEQOTQNIPOQRTVYLPGLERTQNNATT	528
Qy	538	EDLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSQIIESE	597
Db	529	EELHVITLKGAGRAQVRVLHWESGKPEDINNQLRYSYDNLIGSSQLQSDSQIIESEE	588
Qy	598	YYPGGTALWAARNGTEASYKTIYRSGKERDATGLYYGYRYQYQWIGWLSSDPAGTID	657
Db	589	YYPEGGTALWAARNGTEASYKTIYRSGKERDVTGLYYGYRYQYQWAGRWLGADPAGTID	648
Qy	658	GLNLYRMVRNPNVTLDPDGLMPTTIAERIAALKKNKVTDSPSANATNVAL-----N	710
Db	649	GLNLYRMVRNPNVTFQDVOGL-----SPANRTEALIKQGSFTG	687
Qy	711	IRPPVAPKPSLPK-----ASTSSQP-----TTHPIGAANIKP-----TTSGSSIVAPL	753
Db	688	MEEAVYKQWAKPQTFKQRATAAQTEQEAHESLTNPN--SVDISPIKNYTTDSSQINAAI	745
Qy	754	SPVGNKSTSEISLPSAQAQSSSTSTSNLQKKSFTLYRADNKRSPFEGFKAWTP	813
Db	746	RE--NRITPAV---ESLDATLSSLDQRQMRVTRYVMTYVDNS-----TP	784
Qy	814	LDTKWARQFASIFIGQKDTSNLPKETVKNISWGAQPKLKOLSNYIKYTKDKSTVWVSTA	873
Db	785	SPWHSPOEGNSINVGDIVSDNAYLSTSAH-----RGLFNPHVKKETSETRYVKWA	834
Qy	874	INTBAGQSGAPLHKIDMDLYEFAID-----GQKLNPLPEGRTKNMVPSLLIDTPQ	925
Db	835	PLTNAGVNVPAASMYNNAGESEQVFKMDLNDRSKLSAEKLUKRVSPQSQOAILLPRETQ	894
Qy	926	IETSSIIALNH	936

895 PE---VVSMMK 902

DB

RESULT 6

ADR21579

ID ADR21579 standard; protein; 938 AA.

XX

AC ADR21579;

XX

DT 04-NOV-2004 (first entry)

XX

DE Photorhabdus Tcc5 toxin.

XX

KW toxin; insect; insecticidal; transgenic; pest control.

XX

OS Photorhabdus luminescens; W14.

XX

PN WO2004067727-A2.

XX

PD 12-AUG-2004.

XX

PF 07-JAN-2004; 2004WO-US000394.

XX

PR 21-JAN-2003; 2003US-0441723P.

XX

PA (DOWC) DOW AGROSCIENCES LLC.

XX

PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;

XX

PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;

XX

DR WPI; 2004-580999/56.

XX

DR N-PSDB; ADR21537.

XX

PT Controlling or inhibiting an insect, useful for pest control, comprises

XX

PT contacting the insect with effective amounts of a Protein A, a Protein B,

XX

PT and a Protein C.

XX

PS Claim 1; SEQ ID NO 57; 368pp; English.

XX

CC The invention relates to a novel method for controlling or inhibiting an

XX

CC insect comprising contacting the insect with effective amounts of a

XX

CC Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C

XX

CC is encoded by a naturally occurring gene or has an amino acid sequence

XX

CC that differs from the product encoded by a naturally occurring gene only

XX

CC by truncation or by conservative amino acid changes. Protein A is a 230-

XX

CC 290 kDa toxin complex insect toxin that is derived from a first taxonomic

XX

CC species, has stand alone insecticidal activity, and has an amino acid

XX

CC sequence at least 40% identical to a sequence selected from XptAlwi,

XX

CC XptA2w1, TcdA, TcdA2, TcdA4, and TcBa. Protein B is a 130-180 kDa toxin

XX

CC complex potentiator having an amino acid sequence at least 40% identical

XX

CC to a sequence selected from TcdB1, TcdB2, TcdC, XptC1w1, XptC1xb,

XX

CC PptB1(orfs), or SepB. Protein C is a 90-120 kDa toxin complex potentiator

XX

CC having an amino acid sequence at least 35% identical to a sequence

XX

CC selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1w1, XptC1xb, PptC1

XX

CC (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic

XX

CC plant or plant cell that produces a Protein A, a Protein B, and a Protein

XX

CC C. The method is useful for pest control. The present sequence represents

XX

CC Photorhabdus luminescens Tcc5 toxin.

XX

SQ Sequence 938 AA;

Query Match 57.5%; Score 2879; DB 8; Length 938;

Best Local Similarity 60.9%; Pred. NO. 3.7e-191;

Matches 591; Conservative 90; Mismatches 186; Indels 104; Gaps 16;

Qy 1 MKNDPKLYQKTFVSVVDNRGLIIRNIDFHRITANGDPDTRTHQYDIGHLHNSIDP 60

DB 1 MENIDPKLYHHTPTVSVHDNRGLAIRNISPFRITAEANTDTRTHQYNAGGYLHNSIDP 60

Qy 61 RLYEAKQNTNITKNFLKQYDITGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120

DB 61 RLYDAKQTNNAVQPNFIWRHNLITGNILRTESVDAGRTITLNDIEGRPLLTINAAAGVRQNH 120

Qy 121 QYETSLPGRLLSVAEQ-TPEEKTSRITERLIWAGNTEAEKDHNLAGQCVRHYDTAGVTR 179

DB 121 RYEDNTLPGRLLAISEQQAEEKI---TERLIWAGNTPQEKDHNLAGQCVRHYDTAGLTQ 177

Qy 180 LESLSLTGTVLSQSSQLLIDTQEBANWTGDNETWQNMADDIYTTLTSTFDATGALLTQTD 239

DB 178 LNSLALTGAVLSSQQLLTDNQADMTGEDQSLWQQLSDVYITQSTNTDATGALLTQTD 237

Qy 240 AKGNIQRLAYDVAGOLNGSWLTLKGOTEQVLIKSLTYSAAQOKLREBHGNDVITEYSYEP 299

DB 238 AKGNIQRLAYDVAGOLKGSWLTGQAEQVLIKSLTYSAAQOKUREHGNGIYVTEYSYEP 297

Qy 300 ETORLIGIKTRRPSDTKVLQDLRYEYDPVGNVISIRNDAEATRFHWHKQVMPENTYTYDS 359

DB 298 ETQRLIGITTRPSDAKVLQDLRYQYDPVGNVISIRNDAEATRFHWHKQVAPENSYYTDS 357

Qy 360 LYQLISATGREMANIGQSHQFPSPALPSDNNTVNTYTRTYTYDRGNGLTKIQHSSPATQ 419

DB 358 LYQLISATGREMANIGQSNQLPSPALPSDNNTYNTYTRTYTYDRGNGLTKIQHSSPAAQ 417

Qy 420 NNYTNTITVNSNRNRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNWNTRGELQOVTL 479

DB 418 NNYTDTITVNSNRNRAVLSTLTADPTQVDALFDAGGHQNTLISGQNLNWNTRGELKQ--- 474

Qy 480 VKRDKGANDD--REWYRYSYSGDRMLKINEQQAASNNAAQTQRTVYLPNLELTLQNTATT 537

DB 475 -----ANNAGNENWYRYSNGIRQLKNEQQTQNIPOQQRVTVLPGLEIRTTQNNATTT 528

Qy 538 EDLOVITVGBAGRAQVRLVHWSGKPEDINNNQIRSYDNLIIGSSQLELSEGGIISBEE 597

DB 529 EELHVTITLKGAGRAQVRLVHWSGKPEDINNNQIRSYDNLIIGSSQLELSDGQIISBEE 588

Qy 598 YYPYGGTALWAARNQTEASVYKTRYSCKERDATGLYYGYRYQYQWIGRWLSSDPAGTID 657

DB 589 YYPFGGTALWAARNQTEASVYKTRYSCKERDVTGLYYGYRYQYQWIGRWLGDAGTID 648

Qy 746 RE--NRITPAV---ESLDATLSLQDRQMTYRYRVMTYVDS-----TP 784

DB 814 LDTKWARQFASIFIGQKDTSNLPKETVKNISTWGAKPKLKDLSNYIKYTKDKSTVWVSTA 873

DB 785 SPWHSPOEGNSINVGDIVSDNAYLSTSAH-----RGFLNFVHKKETSETRYVQWA 834

Qy 874 INTRAGQSSGAPLHKIDMDIYEFAD-----GQKLNPLPEGRTKMVPFSLLDTPQ 925

DB 835 PLTNAGVNPVPAASMYNNAGEQVFMDLNDRSKSLAEKLLKRVSGSQSQAEILLPRETQ 894

Qy 926 IETSSIIALNH 936

DB 895 FE---VVSMMK 902

RESULT 7

ADP18626

ID ADP18626 standard; protein; 949 AA.

XX

AC ADP18626;

XX

DT 12-AUG-2004 (first entry)

XX

DE Photorhabdus luminescens W-14 tccC4 protein SEQ ID NO:14.

XX

KW tcd; Photorhabdus luminescens W-14; transgenic plant; Toxin A;

XX

KW orally active insect toxin; insect toxin; tccC4.

Photorhabdus luminescens.
MO2004044217-A2.
27-MAY-2004.
12-NOV-2003; 2003WO-IB005553.
12-NOV-2002; 2002US-0425672P.
(UYBA-) UNIV BATH.
French-Constant RH, Waterfield NR;
N-PSDB; ADP18625.
New isolated Photorhabdus luminescens nucleic acids, useful for
expressing orally active insect toxin or for generating transgenic plants
with enhanced resistance to insects.
Claim 1; SEQ ID NO 14; 118pp; English.
The present invention describes DNA sequences from the tcd genomic region
of Photorhabdus luminescens W-14. Also described: (1) a transgenic
monocot or dicot cell having a genome comprising a nucleic acid sequence
that encodes SEQ ID NO:4, 6, 8, 10, 12, 14, or 16; (2) a transgenic plant
with a genome comprising a nucleic acid sequence that encodes SEQ ID
NO:4, 6, 8, 10, 12, 14, or 16; (3) a seed of the transgenic plant; (4) a
progeny of seed; (5) a method of producing toxin A of Photorhabdus
luminescens W-14 in a heterologous host; and (6) a method of producing an
orally active insect toxin. The nucleotide sequences are useful for
heterologous expression of orally active insect toxin. They can also be
used for generating transgenic plants with enhanced resistance to
insects. The present sequence represents Photorhabdus luminescens W-14
tccC4, which is used in the exemplification of the present invention.

Query Match 57.0%; Score 2850.5; DB 8; Length 949;
Best Local Similarity 60.9%; Pred. No. 3.7e-189;
Matches 583; Conservative 100; Mismatches 207; Indels 67; Gaps 17;

QY 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDPHRTTANGDDPTRIIRHOYDIHGLNQSIDP 60
DB 1 MKNIDPKLYQKTPVSVYDNRGLIIRNIDPHRDVAGDDPTRIIRHOYDTRGHLQSIDP 60
QY 61 RLYEAKQTNNIKPNFLQVLDLGNPLCTESIDAGRTVTLNDIEGRPLLTVTATGVIQTR 120
DB 61 RLYDAKQTNNINPNFLQVNLGTDLRTESVDAGRTVALNDIEGRQVLIVTATGAIQTR 120
QY 121 QYETSSLPGRLLSVARQTPPEKTSRTERLIWAGNTAEKDHNLACQVRHYDRAGVTPL 180
DB 121 QYEAANTLPGRLLSVARQTPPEKTSRTERLIWAGNTAEKDHNLACQVRHYDRAGVTPL 180
QY 181 ESLSLGTGVLQSSQLLIDTQANWTGDMNETVQWNLADIVTTLSTFDATGALLTQIDA 240
DB 181 ESLSLTENILSQSRQLLADQAEADWTGNDDETTLQWKLINSETTSTFDATGALLTQIDA 240
QY 241 KGNIORLAYDVAGQLNGSLTLKGQTEQVVIKSLTYSAGQKLRBEHGNVDITEYSPEPE 300
DB 241 KGNMQLAYNVAGQLQSSWLTLLKQSEQVIVKSLTYSAGQKLRBEHGNVDITEYSPEPE 300
QY 301 TORLIGIKTRRPSDTKVQLDLRYEYDPVGNVISIRNDAEATFRWQKMPENTVYDSSL 360
DB 301 TRLRIGITRRQSDSKVLQDLRYEYDHPVGNIIISVRNDAEATFRWQKMPENTVYDSSL 360
QY 361 YQLISATGEMANIGQSHQFSP--ALPSDNNVTNTYTRTYDRGGLTKIOHSSPAT 418
DB 361 YQLISATGEMANIGQSNLPSPILPTDENSNTYTRTYDRGGLTKIOHSSPAT 420
QY 419 ONNYTTNITVSNRRAVLSTLTDEPAVDALFDAGGHONTILISGNLNNWTRGLOQVT 478

Db 421 QNNYTTDITVSNRRAVLSSLTSDPTQVEALFDAGGHQTKLLPGQELSNWTRGELQKVT 480
QY 479 LVKRDKGANDREWRYSGDGRMLKINBQOASNNAAQTQRTVYLPNLELRLTONSTATTTE 538
DB 481 PVSRESAS--DREWRYGNDGMRRLKVSEQOTGNSQOQRTVYLPNLELRLTQNGTTTSE 538
QY 539 DLQVITVGBAGRAQVRVLHWESKPEDIDNNQLRYSYDNLIGSSQLELSEGOIISSEY 598
DB 539 DLHAITVGAAGHAQVRVLHWETTPPAGINNQLRYSYDNLIGSSQLELSEGOIISSEY 598
QY 599 YPYGGTALWAARNOTRASYKTIYRSGKERDATGLYYGYRYQYOPWIGRWLSSDPACTIDG 658
DB 599 YPYGGTALWAARNQIEASYKILYRSGKERDATGLYYGYRYQYOPWIGRWLSSDPACTIDG 658
QY 659 LNLRYMVRNPNVTLDPDGLMPTIAERIAALKKN-----KVTDSPSPANATNVAIN 710
DB 659 LNLRYMVRNPNSTLVDISGLAPT-----KYNIPGDFDVEIDQKSKLKPTRLIRIK 710
QY 711 IR-----PPV-----APKPSL--PKASTSQPTTHPIGAANIK---PTTSGSSIVAPLSPV 756
DB 711 DEFLHYGPVDKLEBKKGGLNVPPEELFDRGSPSENGSVTLTFFKXDLPISCISNTEYTL 770
QY 757 GNSKTSISLSPESAQSSSTSTTNLQKKSFTLYRADNRSFEEMQSK-PPEGFKAMTPLD 815
DB 771 YNHETK-PPPYENEAATVGADLGV-IMSVFEGNKSIGNASDEDLKEHLPLGKSTWMDKT 828
QY 816 TKMARQFASIFQKQDTSNL-----PKETVKNISTGAKPKLKDLSNY 858
DB 829 LPDLQK--GLMIAEKIKSGKGYPFHFGAAIAVYGEDKVAASILTDLSEPK-RDEGEY 885
QY 859 IKYTKDKSTVWVSTAINTEAGGSSGAPLHKIDMDLYEFA-IDGQKLNPLPEGRTKN 914
DB 886 LQSTRKVSAMFI-TNVNEFRG-----HDYPKSKYSICLVTAERQPVISKRRAN 933

RESULT 8
ADR21544
ID . ADR21544 standard; protein; 949 AA.
XX AC ADR21544;
XX DX 04-NOV-2004 (first entry)
XX DE Photorhabdus TccC4 toxin SEQ ID NO:64.
XX KW toxin; insect; insecticidal; transgenic; pest control.
XX OS Photorhabdus luminescens.
XX PN WO2004067727-A2.
XX PD 12-AUG-2004.
XX PF 07-JAN-2004; 2004WO-US000394.
XX PR 21-JAN-2003; 2003US-0441723P.
XX PA (DOWC) DOW AGROSCIENCES LLC.
XX PI Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX PI Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
XX DR WPI; 2004-580999/56.
XX CC Controlling or inhibiting an insect, useful for pest control, comprises
XX CC contacting the insect with effective amounts of a Protein A, a Protein B,
XX CC and a Protein C.
XX PS Claim 1; SEQ ID NO 64; 368pp; English.
XX CC The invention relates to a novel method for controlling or inhibiting an
XX CC insect comprising contacting the insect with effective amounts of a

Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C is encoded by a naturally occurring gene or has an amino acid sequence that differs from the product encoded by a naturally occurring gene only by truncation or by conservative amino acid changes. Protein A is a 230-290 kDa toxin complex insect toxin that is derived from a first taxonomic species, has stand alone insecticidal activity, and has an amino acid sequence at least 40% identical to a sequence selected from xPcAlwi, xPcA2wi, TcdA, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin complex potentiator having an amino acid sequence at least 40% identical to a sequence selected from TcdB1, TcdB2, TcdB3, TcdC, xPcC1wi, xPcB1xb, xPcB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator having an amino acid sequence at least 35% identical to a sequence selected from TccC1, TccC2, TccC3, TccC4, TccC5, xPcC1wi, xPcC1xb, PptC1 (orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic plant or plant cell that produces a Protein A, a Protein B, and a Protein C. The method is useful for pest control. The present sequence represents *Photobacterium luminescens* TccC4 toxin.

XX Sequence 949 AA;

Query Match 57.0%; Score 2850.5; DB 8; Length 949;
 Best Local Similarity 60.9%; Pred. NO. 3.7e-189;
 Matches 583; Conservative 100; Mismatches 207; Indels 67; Gaps 17;

Qy 1 MKNIDPKLYQKPTVSVYVYDNRGLIIRNIDFHTTANGDPDTRITRHOYDIHGHILNQSIDP 60
 Db 1 MKNIDPKLYQHTPTVYVYDNRGLIIRNIDFHRVAGGDTDTTRITRHOYDTTGHLSQSIDP 60
 Qy 61 RLYEAKOTNNIKKNFLWQYDLTGPNLCESIDAGRTVTLNDIEGRPLLTATGVTQTR 120
 Db 61 RLYDAKQNNSTNPNFLWQYDLTGDTLTESVDAGRTVALNDIEGRVLIIVTATGQTR 120
 Qy 121 QYETSSLPGRLLSVAEOPPEKTSRITERLIWAGNTEAKDHNLGACVRHYDTAGVTRL 180
 Db 121 QYEANTLPGRLLSVEQAPGEQTPRVTEHFIWAGNTEAKDHNLGACVRYHYDTAGVTL 180
 Qy 181 ELSLSITGVLSQSOLLIDTOEANNWGDNETWQNMADDIYTLTSTFDATGALLTQIDA 240
 Db 181 ELSLSITENILSOSQLLADGQADWGTNDFTLWQKLNSETTYTQSTFDATGALLTQIDA 240
 Qy 241 KGNTRQALYVAGQINGSLWLTGKQTEQVVIKSLTYSAAQKLEEHGNDVITEYSYEPE 300
 Db 241 KGNWQALYVAGQIQGSLWLTGKQSEQVIKSLTYSAAQKLEEHGNGVITEYSYEPE 300
 Qy 301 TORLIGIKTRPSDTKVLQDLRYEYDVPVGNVISIRNDAEATRFWNNKQMPENTYTDLS 360
 Db 301 TLRLLGTTRRSQSDSKVLQDLRYEYDHPVGNIIISVRNDAEATRFWRNKKIVPENTYTDLS 360
 Qy 361 YOLISATGREMANIGQOOSHOPSP--ALPDSNNVTYNTYTRTYVDRGNLTKIOHSSPAT 418
 Db 361 YOLISATGREMANIGQOQNLPSPPIIPTDENSNTYNTYTRTYVDRGNLTVQIRHSSPA 420
 Qy 419 QNNYTTNITVNSRRAVLSTLTPDPAQVDALFDAGGHQNTLISQNLNMMNRGELQOVT 478
 Db 421 QNNYTTDITVNSRRAVLSSITSDPTQVEALFDAGGHQTKLLPQGLSNNTRGELKQVT 480
 Qy 479 LVKDKKGANDREWTRYSGDGRMKINEQQAASNAQQTQVYLPNLERLTQNSTATTE 538
 Db 481 PVSRESAS--DREWTRYSGDGRMKLVKSEQQTGNSQOQRTVYLPDLRLTQNGTTTSE 538
 Qy 539 DLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIISBEY 598
 Db 539 DLHAIITVGAAGHAQVRVLHWETTPPAGINNQLRYSYDNLIGSSQLELDNAGQIISBEY 598
 Qy 599 YPYGCTALWAARNQTEASYKTIYRYSKGERDATGLYYGYRYQYQPWIGRWLSSDPAGTIDG 658
 Db 599 YPFGCTALWAARNQLEASYKILYRYSKGERDATGLYYGYRYQYQPVWGWLSADPAGTIDG 658
 Qy 659 NNLVYMNPNVTLDDPGLMPTTIAERTAAUKN-----KVTDSPAPSNATVAIN 710
 Db 659 NNLVYMNPNSTLVDISGLAPT-----KYNIPGDFDVEIDQKRSLKPTLIRK 710
 Qy 711 IR----PPV----APKPSL--PKASTSSQPTTHPIGAANTK---PTTSGSSIVAPLSPV 756

Db 711 DEFLHYGPDVKLLLEKKPGLNVPPEELFRGSENGVSTLTKDKLPISCISNTBVTLDIL 770
 Qy 757 GKNKSTSIISLPESAQSSSSTSTTNLQKKSFTLYRANRSPFEEMQSK-FPEGFKAWTPLD 815
 Db 771 YNKHETK-PPPYENEATVGDALGV-IMSVBFGNKSIGNASDEDLKEEHLPLGKSTMDKTD 828
 Qy 816 TKMARQFASIFIGQKDTSNL-----PKETVKNISTWGAKPKLKLDSNY 858
 Db 829 LPDLKQ--GLMIAEKIKSGKAGYFPHFGAAIAVYVGDKKVAASILTDLSEPK-RDEGEY 885
 Qy 859 IKYTKDKSTVWVSTAINTEAGQSSGAPLHKIDMDLYEFA-IDGQKLNPLPEGRTKN 914
 Db 886 LQSTRKVSAMPI-TNVNEPRG-----HDYPSKYSIGLVTAERKQPVISKRRAN 933

RESULT 9
 ABM70226
 ID ABM70226 standard; protein; 939 AA.
 XX
 AC ABM70226;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Photobacterium luminescens protein sequence #3323.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX
 OS Photobacterium luminescens.
 XX
 FN WO200294867-A2.
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-IB003040.
 XX
 PR 07-FEB-2001; 2001FR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 PI Buchrieser C;
 XX
 DR WPI; 2003-148459/14.
 XX
 PT Genomic sequence of Photobacterium luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 2; SEQ ID NO 3323; 1205pp; French.
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photobacterium luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.

CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX
 SQ Sequence 939 AA;

Query Match 56.6%; Score 2833.5; DB 6; Length 939;
 Best Local Similarity 58.8%; Pred. No. 5.5e-188;
 Matches 584; Conservative 107; Mismatches 188; Indels 115; Gaps 18;

QY 1 MKNIDPKLYQKPTVSVVDNGLIRNIDFRTTANGDPDTRITRHOYDIHGHNLQSIDP 60
 DB 1 MKNIDPKLYHPTVSVVDNGLAIRNISFRATAEANTDPRITRHOYNAGYLQSIDP 60

QY 61 RLYEAKQNTNITKPNFLQYDLTGMLCTESIDAGRTVTLANDIEGRPLLTATGVIQTR 120
 DB 61 RLYDAKQTNNAVQPNFQHNLGTNLRTESVDAGRTITLANDIEGRPLVLTISATGVRQNH 120

QY 121 QYETSSLPGRLLSVAEQT-PEKTSRITERLIWAGTEAEKHNLAGOCVHYDTAGVTR 179
 DB 121 LYEDNTLPGRLLAITEQAQTEKT---TERLIWAGTNPQKEYNLVGQCTRHYDTAGLAQ 177

QY 180 LESLSLTGTVLSQSSOLLIIDTQEANWGTNETVWQNLADLIYTLSTFDATGALLTOTD 239
 DB 178 LNSLALTGAVLSQSQOPLVDNQADWGTGDSLWQOKLSSDYTTQNTKDTAGVLLTOTD 237

QY 240 AKGNITQRLAYDVAGQLGSLWTLKQTEQVILKSLITYSAAGQKLEEHNGNDVITEYSYEP 299
 DB 238 AKGNIQRAYDVAGQLKGSWTLKQTEQVILKSLITYSAAGQKLEEHNGNGLITEYSYEP 297

QY 300 ETORLIGIKTRRPSKTVQLDRLRYDPVGNVISIRNDAEATRFWKNQKVPENTYTVDS 359
 DB 298 ETORLIGITRRPSDAKVLQDLRYQYDPVGNVINIRNDAEATRFWRNOKVIPENSYTVDS 357

QY 360 LYQLISATGREMANIQGOHQPPSPALPDNNYTNRTYTYDRGGLTKIOHSSPATQ 419
 DB 358 LYQLISATGREMANIQGNPLPSPALPDNNYTNRTYTYDRGGLTKIOHSSPATQ 417

QY 420 NNYTNTITVSNRSLVLTSTEDPAQVDALFDAGGHQNTLISGQNLNWNTRGELQQVTL 479
 DB 418 NNYTNTITVSNRSLVLTSTEDPTQVDALFDGSGHQTLLSGQVLWTFRGELKQV-- 475

QY 480 VKRDGANDREWYSGDGRMLKINEQOASNAQTORVYVLPNLELRITONSTATTED 539
 DB 476 ---NSSAGNEVHYDSNGTRQLKVNQOQTQIAQQOQRTVYLPGLERLTTHQGSTTTEY 530

QY 540 LQVITVGEAGRAQVRLHWESKPEDINNLQRYSDNLISSQLELSEGGIISEERY 599
 DB 531 LQVITLKGAGRAQVRLHWESKPEDINNNQRYSDNLISSQLELSEGGIISEERY 590

QY 600 PYGGTALWAARNQTEASYKTRYSKGERDATGLYYGYRYTQPTWIGRWLSDPPACTIDGL 659
 DB 591 PFGGTALWAARNQTEASYKTRYSKGERDATGLYYGYRYTQPTWIGRWLSDPPACTIDGL 650

QY 660 NLYRMVRNPNVTLDPDGLMPTIAERIAALKKNKVTDSAPSPANATNVAI-----NIR 712
 DB 651 NLYRMVRNPNVTFQVQGL-----SPANRTEALIKQGSFTGWE 689

QY 713 PPVAPKPSLPKA-----STSSQPTTHPIGAAN-----IKPTTSGSSIVAPLSVGN 758
 DB 690 EAVYKMAKQPTFKQRAIATQTEAHQOLLNPNPGVDTSPDKYTTDSSQI-----N 742

QY 759 KSTSE---ISLPESAQSSSTTSNTLQKSPFLYRANRSPFEEMQSKFPEGKAWTPLD 815
 DB 743 TAIRENRITISVKDLDSLSALQDRIQVRYVWYIDN-----SK-----PSP 786

QY 816 TKMARQFASIFGQKQDQTNLPKPTVKNTSTWGAQKPLKDLNVIKYTKDKSTVWVSTAIN 875
 DB 787 WHSPQEGNSINVDIVSDNAVLSAHS-----RGFLNFVHKKETSETRYVKMAFL 836

QY 876 TEAG-----GSSGNPLHKIDM-DLYEPAIDGQKL--NPLPEGRTKNWPVSLLDLT 923
 DB 837 TNTGVNVAASKYNNENAEKIFKMDLDDSWKSFVEKLIKIRANGPAGQAEILFPR---ET 893

QY 924 POIETSSIIALNHGPNVDNAEISFLTTIPLKNVKP 957
 DB 894 P-----FEVVSRRHQGRD-----TYVLLQDIKP 916

RESULT 10
 ABM70214
 ID ABM70214 standard; protein; 960 AA.

XX AC ABM70214;

XX 20-NOV-2003 (first entry)

XX Photorhabdus luminescens protein sequence #3311.

DE Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.

XX Photorhabdus luminescens.

OS WO200294867-A2.

XX 28-NOV-2002.

XX 07-FEB-2002; 2002WO-IB003040.

XX 07-FEB-2001; 2001FR-00001659.

XX (INSP) INST PASTEUR.

PA (CNRS) CNRS CENT NAT RECH SCI.

XX Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;

PI Buchrieser C;

XX WPI; 2003-148459/14.

XX Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 useful e.g. as therapeutic antimicrobials and agricultural pesticides.

XX Claim 2; SEQ ID NO 3311; 1205pp; French.

PS The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins

XX Sequence 960 AA;

Query Match 56.4%; Score 2821.5; DB 6; Length 960;

Best Local Similarity 67.0%; Pred. No. 3.9e-187;

Matches 550; Conservative 74; Mismatches 133; Indels 65; Gaps 7;

QY 1 MKNIDPKLYQKPTVSVVDNGLIRNIDFRTTANGDPDTRITRHOYDIHGHNLQSIDP 60

```
Db 1 MKNIDPKLYOHTPTISVYDNRGLTIRNIDFHSVAGGDTDTTRITRHQYDVGRHLSQSIDP 60
Qy 61 RLYEAKQTNNTIKNFVWYDLDTCNPLCTESIDAGRTVTLNDIEGRPLLTATATGVTQTR 120
Db 61 RLYDAKOLDNSINFVWYDLDTCNPLCTESIDAGRTVTLNDIEGRPLLTATATGVTQTR 120
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLWAGNTAEKHNLGAGQVRYHDYTAGVTRL 180
Db 121 QYEAUTLPGRLLSVAEQTPPEKTSRITERLWAGNTAEKHNLGAGQVRYHDYTAGVTRL 180
Qy 181 ELSLTGTVLSSQSLIDTQEAANNWGTNETVQWMLADDIYTLSTFDATGALLTQTD 240
Db 181 KSLSTENILSSQSLIDTQEAANNWGTNETVQWMLADDIYTLSTFDATGALLTQTD 240
Qy 241 KGNITQRLAYDVAGQINGSLWTLKQTEQVIKTSYSAAGOKLREHNGNDVITEYSPE 300
Db 241 KSNITQRLAYDVAGQINGSLWTLKQTEQVIKTSYSAAGOKLREHNGNDVITEYSPE 300
Qy 301 TORLIGIKTRRPSPTKVLQDLRYEYDPVGNVISIRNDAEATFRWHNOKVMPENTYTYD 360
Db 301 TORLIGIKTRRPSPTKVLQDLRYEYDPVGNVISIRNDAEATFRWHNOKVMPENTYTYD 360
Qy 361 YQLISATGREMANIGQSHQSPSP-ALPDSNNNTYNTYTRTYDRGNGNLTQIHSPPATQ 419
Db 361 YQLISATGREMANIGQSHQSPSP-ALPDSNNNTYNTYTRTYDRGNGNLTQIHSPPATQ 419
Qy 420 NNYTNTITVNSRNVAVLSTEDPAQVDALFDAGGHONTLSGONLWNTNRGLOQVTL 479
Db 420 NNYTNTITVNSRNVAVLSTEDPAQVDALFDAGGHONTLSGONLWNTNRGLOQVTL 479
Qy 480 VIKDKGANDREWYSGDGRMLKINEQOASNAQOTVRYLNLLELRLTONSTATED 539
Db 480 VIKDKGANDREWYSGDGRMLKINEQOASNAQOTVRYLNLLELRLTONSTATED 539
Qy 540 LQVITVAGRAQVRVHLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSEGOIISEBEY 599
Db 540 LQVITVAGRAQVRVHLHWSGKPEDIDNNQLRYSYDNLIGSSQLELDSEGOIISEBEY 599
Qy 599 LHVITVGAAGHAQVRLHWATTPAGISNNQLRYSYDNLIGSSQLELDSEGOIISEBEY 658
Db 599 LHVITVGAAGHAQVRLHWATTPAGISNNQLRYSYDNLIGSSQLELDSEGOIISEBEY 658
Qy 660 NLYRVRNPNVTLADPGLMPTIAERIAALKKNKVTDSAPSPANATVAINIRPPVAPKP 719
Db 660 NLYRVRNPNVTLADPGLMPTIAERIAALKKNKVTDSAPSPANATVAINIRPPVAPKP 719
Qy 699 NLYRVRNPNVTLADPGLMPTIAERIAALKKNKVTDSAPSPANATVAINIRPPVAPKP 757
Db 699 NLYRVRNPNVTLADPGLMPTIAERIAALKKNKVTDSAPSPANATVAINIRPPVAPKP 757
Qy 720 SLPKASTSSQ-----PTTH---PIGAANIKPTT-----SGSIVAPLSPVG 757
Db 720 SLPKASTSSQ-----PTTH---PIGAANIKPTT-----SGSIVAPLSPVG 757
Qy 758 NKSTSEISLPSAQSSTSTSTNLQKKSFTLYRADNRSFE 798
Db 758 NKSTSEISLPSAQSSTSTSTNLQKKSFTLYRADNRSFE 798
Qy 747 YSRSLKVAASWAASGSGRVVKFRICK-----NNRSFD 779
Db 747 YSRSLKVAASWAASGSGRVVKFRICK-----NNRSFD 779
```

RESULT 11

ABM69154
ID ABM69154 standard; protein; 966 AA.

AC ABM69154;

DT 20-NOV-2003 (first entry)

DE Photorhabdus luminescens protein sequence #2251.

KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;

KW whooping cough.

OS Photorhabdus luminescens.

XX WO200294867-A2.

XX PD 28-NOV-2002.
XX PF 07-FEB-2002; 2002WO-IB003040.
XX PR 07-FEB-2001; 2001FR-00001659.
XX PA (INSP) INST PASTEUR.
XX PA (CNRS) CNRS CENT NAT RECH SCI.
XX PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
XX PI Buchrieser C;
XX DR WPI; 2003-148459/14.
XX PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
XX PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX PS Claim 2; SEQ ID NO 2251; 1205pp; French.

CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins

Sequence 966 AA;

Query Match 55.3%; Score 2768.5; DB 6; Length 966;
Best Local Similarity 64.0%; Pred. No. 1.9e-183;
Matches 541; Conservative 96; Mismatches 165; Indels 43; Gaps 5;

Qy 1 MKNIDPKLYOHTPTISVYDNRGLTIRNIDFHSVAGGDTDTTRITRHQYDVGRHLSQSIDP 60
Db 1 MNFPDKLYOHTPTITVHDNRGLDVREIHYHRAEAKNSDIRTHQYDVGRHLSQSIDP 60
Qy 61 RLYEAKQTNNTIKNFVWYDLDTCNPLCTESIDAGRTVTLNDIEGRPLLTATATGVTQTR 120
Db 61 RLYDAWKQSSVKNFVWYDLDTCNPLCTESIDAGRTVTLNDIEGRPLLTATATGVTQTR 120
Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLWAGNTAEKHNLGAGQVRYHDYTAGVTRL 180
Db 121 QYEDNSLPGRLLSVTEQEPDERASHITERLWAGNTAEKHNLGAGQVRYHDYTAGVTRL 180
Qy 181 ELSLTGTVLSSQSLIDTQEAANNWGTNETVQWMLADDIYTLSTFDATGALLTQTD 240
Db 181 NLSLTGALSSQSLIDTQEAANNWGTNETVQWMLADDIYTLSTFDATGALLTQTD 240
Qy 241 KGNITQRLAYDVAGQINGSLWTLKQTEQVIKTSYSAAGOKLREHNGNDVITEYSPE 300
Db 241 KGNITQRLAYDVAGQINGSLWTLKQTEQVIKTSYSAAGOKLREHNGNDVITEYSPE 300
Qy 301 TORLIGIKTRRPSPTKVLQDLRYEYDPVGNVISIRNDAEATFRWHNOKVMPENTYTYD 360
Db 301 TORLIGIKTRRPSPTKVLQDLRYEYDPVGNVISIRNDAEATFRWHNOKVMPENTYTYD 360
Qy 361 YQLISATGREMANIGQSHQSPSP-ALPDSNNNTYNTYTRTYDRGNGNLTQIHSPPATQ 420

Db 361 YQLISATGEMANTGQONLPESTALPSDNNNTYTRYAYDRGGLNLTQIRHSSPASQN 420
 Qy 421 NYTNTITVSNSRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTTREGELQOQVTLV 480
 Db 421 NYTIDITVSNSRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTTREGELQOQVTLV 480
 Qy 481 KRDK--GANDREWRYSGDGRMLKINEQQAASNNAAQORVYVLPNLELRTONSTATTE 538
 Db 481 KRDLNLTDRSDIEWRYGNDGWRLLKVSQEQMONILQORRVVYLSGLELRTQNGDITKE 540
 Qy 539 DLOVITVGEAGRAQVRVLHWSGKPEDIDNNLRYSYDNLIGSSQLELDSGQIISREY 598
 Db 541 ELQIITVGEAGRAQVRVLHWSGKPEINNNQVRYSYGNLIGSSQLELDSGQIISREY 600
 Qy 599 YPYGCTALWAARNQTEASYKIRYSGKRDATGLYYGYRYQYQWIGRWLSDDPAGTIDG 658
 Db 601 YPYGCTALWAARNQTEASYKIRYSGKRDATGLYYGYRYQYQWIGRWLSDDPAGTIDG 660
 Qy 659 LNLRYMVRNPNVTLDPGLMPTI-----AER-----IAALKK- 691
 Db 661 LNLRYMVRNPNVTLDPGLMPTI-----AER-----IAALKK- 720
 Qy 692 -----NKVTDAPSANATVAINIRPPVAPKPSLPKASTSQPTTHPIGAANIKPTT 744
 Db 721 FAATSWGKVKSDTKTILSASAEYGNLKHKKQMEYNNKKLNKNETSPDGKFKYAKTKLQDQA 780
 Qy 745 SGSSIVAPLSPVGNKSTSEISLPSAQSSSSSTSTNLQKKSFTLYRADNRSFEEMQSKF 804
 Db 781 AHAGIAFNHLPISONE-----KEGWTFNKDFVNDKDKSLGKGLKGLPFGVELLPDN 833
 Qy 805 PEGFK 809
 Db 834 PEHYK 838

RESULT 12
 ABM69055
 ID ABM69055 standard; protein; 1044 AA.
 XX AC ABM69055;
 XX DT 20-NOV-2003 (first entry)
 XX DE Photothabidus luminescens protein sequence #2152.
 XX KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough.
 XX OS Photothabidus luminescens.
 XX FN W0200294867-A2.
 XX PD 28-NOV-2002.
 XX PF 07-FEB-2002; 2002WO-IB003040.
 XX PR 07-FEB-2001; 2001FR-00001659.
 XX PA (INSP) INST PASTEUR.
 XX PA (CNRS) CNRS CENT NAT RECH SCI.
 XX PA Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
 PI Buchrieser C;
 XX WPI; 2003-148459/14.
 XX Genomic sequence of Photothabidus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX Claim 2; SEQ ID NO 2152; 1205pp; French.

XX The invention relates to the isolation of genes and their encoded
 CC proteins from Photothabidus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The
 CC genes, proteins, vectors containing the genes and Ab are also useful
 CC therapeutically (to treat microbial infection by bacteria or fungi that
 CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
 CC biopesticides. Other uses of the genes and the proteins are as virulence
 CC factors and for identifying targets of human diseases for which P.
 CC luminescens is a model (particularly plague and whooping cough). This
 CC sequence represents one of the isolated P. luminescens proteins
 XX SQ Sequence 1044 AA;
 Query Match 54.8%; Score 2744; DB 6; Length 1044;
 Best Local Similarity 74.7%; Pred. No. 1.1e-181;
 Matches 513; Conservative 78; Mismatches 88; Indels 8; Gaps 3;
 Qy 1 MKNTDPKLYQKTPVSYVDNRGLIIRNIDFRTTANGDPDTRITRHOYDIGHLHNSIDP 60
 Db 1 MSTPTALYTQTPVSYVDNRGLIIRNIDFRTTANGDPDTRITRHOYDIGHLHNSIDP 60
 Qy 61 RLVEAKQNTNNTIKENFLWOYDLTGNCPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120
 Db 61 RLVDKQADNSVKFNFWQYDLGHALTESVDAGRTVALNDIEGRPVMWATGVRQTR 120
 Qy 121 QYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTEAKDHNLGAGQCVRYHDTAGVTRL 180
 Db 121 RYEGNTLPGRLLSVSEQVDFQETSQVTERFIAEWNTTAEKYNLSGLCIRHYDTAGVTRL 180
 Qy 181 ELSLITGTVLSQSSQLLIDTQEAANTGNETWQNLADDIYTLTSTFDATGALLTQD 240
 Db 181 MSQSLAGAILSQSHQWLABGQEAANNMGDETFWQMLASVEYTTQSTTNAIGALLTQD 240
 Qy 241 KGNLQRLAYDVAGOLNGSWLTKGTEQVIKSLITYSAAGOKLREHNGDVITYSYRPE 300
 Db 241 KGNLQRLAYDVAGOLNGSWLTKGTEQVIKSLITYSAAGOKLREHNGDVITYSYRPE 300
 Qy 301 TQRLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWHNQKVPENTY 356
 Db 301 TQRLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWHNQKVPENTY 356
 Qy 357 YDSLYQLISATGEMANTGQONLPESTALPSDNNNTYTRYAYDRGGLNLTQIRHSSPASQN 414
 Db 361 YDSLYQLISATGEMANTGQONLPESTALPSDNNNTYTRYAYDRGGLNLTQIRHSSPASQN 420
 Qy 415 SPATONNTYTNITVSNSRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTTREGEL 474
 Db 421 SPATONNTYTNITVSNSRAVLSTLTEDPAQVDALPDAGGHQNTLISGQNLNNTTREGEL 480
 Qy 475 QQVTLVKRDKGANDREWRYSGDGRMLKINEQQAASNNAAQORVYVLPNLELRTONSTATTE 534
 Db 481 QRVTPVSRN-----SSDSEWRYSSDGVRLLYKVSQEQTNGSTQVRVYVLPGLRLITGVAD 538
 Qy 535 ATTSDQLQVITVGEAGRAQVRVLHWSGKPEDIDNNLRYSYDNLIGSSQLELDSGQIISREY 594
 Db 539 KTTQNLQVITVGEAGRAQVRVLHWSGKPEDIDNNLRYSYDNLIGSSQLELDSGQIISREY 598
 Qy 595 EEEYYPYGGTALWAARNQTEASYKIRYSGKRDATGLYYGYRYQYQWIGRWLSDDPAG 654
 Db 599 QEEYYPYGGTALWAARNQTEASYKIRYSGKRDATGLYYGYRYQYQWIGRWLSDDPAG 658

PR	06-NOV-1996;	96US-00743699.	
PR	06-NOV-1996;	96WO-US018003.	
XX	(DOWC)	DOWELANCO.	
PA	(WISC)	WISCONSIN ALUMNI RES FOUND.	
XX			
PI	Ensign JC, Bowen DJ, Petell J, Fatig R, Schoonover S;		
PI	Ffrench-Constant RH, Rocheleau TA, Blackburn MB, Hey TD, Merlo DJ;		
PI	Orr GL, Roberts JL, Strickland JA, Guo L, Ciche TA, Sukhapinda K;		
XX			
DR	WPI, 1998-179427/16.		
DR	N-PSDB; AAV29927.		
PT	Isolated toxins from Photorhabdus luminescens strains - useful for		
PT	control of insect pests.		
XX			
PS	Claim 34; Page 286-288; 321pp; English.		
XX			
CC	The present sequence represents a protein named TccC of the bacterium		
CC	Photorhabdus luminescens (W-14). This is a symbiotic bacterium of the		
CC	nematodes of the Heterorhabditis genus. The bacterium has at least 4		
CC	distinct genomic regions, tca, tcb, tcc, and tcd. Peptide products are		
CC	produced from these regions that are associated with insecticidal		
CC	activity. The native toxins are secreted proteins. The proteins are toxic		
CC	to insects upon exposure and especially when ingested. The nucleic acid		
CC	sequence can be used to produce transgenic plants, baculoviruses or		
CC	microbial hosts for toxin production. They can be used to control insects		
CC	pests from the Lepidoptera, Coleoptera, Hymenoptera, Diptera,		
CC	Diclyoptera, Acarina or Homoptera orders, especially the Southern or		
CC	Western corn rootworm, Colorado potato beetle, mealworm, boll weevil,		
CC	turf grub, beetle armyworm, black cutworm, cabbage looper, codling moth,		
CC	corn earworm, European corn borer or tobacco hornworm or budworm		
XX			
SQ	Sequence 1043 AA;		
Query Match	54.7%; Score 2740; DB 2; Length 1043;		
Best Local Similarity	74.7%; Pred. No. 2.1e-181;		
Matches 513; Conservative 78; Mismatches 88; Indels 8; Gaps 3;			
QY	1 MKNIDPKLYQKTPVVDVNDGLIIRNIDFHTTANGPDPTRTTRHOYDIHGLNOSIDP 60		
Db	1 MSPSETTYTQPTVSVLDNRGLSIRDIGFRIVIGGDTDRVTHKHQYDARGHLNYSIDP 60		
QY	61 RLYBAKQTNNTIKNPLQYDITGNPLCTESIDAGRTVTLNDIEGRPLLTATGVIQTR 120		
Db	61 RLYDAKQADNSKPNFVQHDLAGHALATESVDAGRTVALNDIEGRSVMTNATGVQTR 120		
QY	121 QYETSSLPGRLLSVAQTPPEKTSRITERLIWAGNTEAKOHNLAGQCVRYHYDTAGVTRL 180		
Db	121 RYEGNTLPGRLLSVEQVFNQESAKVTERFIWAGNTTSEKYNLSGLCIRHYDTAGVTRL 180		
QY	181 ELSLSITGVLSQSSOLLIDTQEAQNTGNETVQWNLADDLYTTLSTFDATGALLTQTD 240		
Db	181 MSQSLAGAMLQSQHLLAQGEQANWSGDETVQGMASEVYTTQSTTNAIGALLTQTD 240		
QY	241 KGNITQRLAYDAGQLNGSLWTLKGTEQVVIKSLIYSAAGKLRHEHNDVITYEYRPE 300		
Db	241 KGNITQRLAYDAGQLNGSLWTLKGTEQVVIKSLIYSAAGKLRHEHNDVITYEYRPE 300		
QY	301 TORLIGIKTRR-----PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWNRKQVPENYT 356		
Db	301 TORLIGITTRAEQSQSQARVLQDLRYEYDPVGNVISIRNDAEATRFWRNRKQVPENYR 360		
QY	357 YDSLYQLYSATCREMANIGQSHOPSAL--PSDNNYTYNTRYTYDRCGNLTQIHS 414		
Db	361 YDSLYQLYSATCREMANIGQSHOPSAL--PSDNNYTYNTRYTYDRCGNLTQIHS 420		
QY	415 SPATQNTYNTITVSNRNRAVLSTLTEDPAQVDALFDAGGHQNTLISQGNLNNWTRGEL 474		
Db	421 SPATQNTYNTITVSNRNRAVLSTLTEDPAQVDALFDAGGHQNTLISQGNLNNWTRGEL 480		
QY	475 QQVTLVKRDKGANDREWYRSGDGRMLKINEQQAASNAQTRVYLYLPNLELRITQNST 534		

Db	481 QRVTPEVSRN--SSDSEWYRYSSDGMRLLLKVSBOQTGNSTQVRVYLYLPGLERTGTGAD 538
QY	535 ATTEDLOVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSGQIIS 594
Db	539 KTTELOVITVGEAGRAQVRVLHWESGKPTDIDNNQVRYSYDNLIGSSQLELDSGQILS 598
QY	595 BEEYYPYGGTALWAARNOQTEASYKTIYRSGKRDATGLYYGYRYQYQWIGRWLSSDPAG 654
Db	599 QEEYYPYGGTALWAARNOQTEASYKFIYRSGKRDATGLYYGYRYQYQWIGRWLSSDPAG 658
QY	655 TIDGLNLYRMVRNPNVTLDPDGLMPT 681
Db	659 TVDGLNLYRMVRNPNITLTDHGLAPS 685
RESULT 15	
ADR21577	
ID	ADR21577 standard; protein; 1043 AA.
XX	
AC	ADR21577;
XX	
DT	04-NOV-2004 (first entry)
XX	
DE	Photorhabdus TccC1 toxin.
XX	
KW	toxin; insect; insecticidal; transgenic; pest control.
XX	
OS	Photorhabdus luminescens.
XX	
FN	WO2004067727-A2.
XX	
PD	12-AUG-2004.
XX	
PF	07-JAN-2004; 2004WO-US000394.
XX	
PR	21-JAN-2003; 2003US-0441723P.
XX	
PA	(DOWC) DOW AGROSCIENCES LLC.
XX	
PI	Hey TD, Schleper AD, Bevan SA, Bintrim SB, Mitchell JC, Li ZS;
XX	Ni W, Zhu B, Merlo DJ, Apel-Birkhold PC;
DR	WPI; 2004-580999/56.
XX	N-PSDB; ADR21505.
PT	Controlling or inhibiting an insect, useful for pest control, comprises
PT	contacting the insect with effective amounts of a Protein A, a Protein B,
XX	and a Protein C.
PS	Claim 1; SEQ ID NO 25; 368pp; English.
XX	
CC	The invention relates to a novel method for controlling or inhibiting an
CC	insect comprising contacting the insect with effective amounts of a
CC	Protein A, a Protein B, and a Protein C. Each of the Proteins A, B, and C
CC	is encoded by a naturally occurring gene or has an amino acid sequence
CC	that differs from the product encoded by a naturally occurring gene only
CC	by truncation or by conservative amino acid changes. Protein A is a 230-
CC	290 kDa toxin complex insect toxin that is derived from a first taxonomic
CC	species, has stand alone insecticidal activity, and has an amino acid
CC	sequence at least 40% identical to a sequence selected from XptAlwI,
CC	XptA2wI, TcdA, TcdA2, TcdA4, and TcdB. Protein B is a 130-180 kDa toxin
CC	complex potentiator having an amino acid sequence at least 40% identical
CC	to a sequence selected from TcdB1, TcdB2, TcdC, XptC1wI, XptB1x,
CC	PptB1(orf5), or SepB. Protein C is a 90-120 kDa toxin complex potentiator
CC	having an amino acid sequence at least 35% identical to a sequence
CC	selected from TccC1, TccC2, TccC3, TccC4, TccC5, XptB1wI, XptC1xb, PptC1
CC	(orf 6 long), PptC1 (orf 6 short), and SepC. Also claimed is a transgenic
CC	plant or plant cell that produces a Protein A, a Protein B, and a Protein
CC	C. The method is useful for pest control. The present sequence represents
XX	Photorhabdus luminescens TccC1 toxin.
SQ	Sequence 1043 AA;

Query Match		54.7%;	Score 2740;	DB 8;	Length 1043;
Best Local Similarity		74.7%;	Pred. No. 2.1e-181;		
Matches 513;		Conservative 78;	Mismatches 98;	Indels 8;	Gaps 3;
Qy	1	MKNIDPKLYQKTPVSVYDNRGLIIRNIDFHRITTTWANGDDPTTRTHQYDIHGHLNQSIDP	60		
Db	1	MSPSETTLYQTPTVSVLDRGLSIRDIGFHRIVIGGDTTRVTRHGYDARGHLNYSIDP	60		
Qy	61	RLYEAKQTNNTIKNFLEWQYDLTGNPLCTESIDAGRTVTTLNDIEGRPLLVTATGVIQTR	120		
Db	61	RLYDAKQADNSVKFNFWQHDLAGHALRTESVDAGRTVALNDIEGRSVMTMNATGVRQTR	120		
Qy	121	OYETSSLPGRLLSVAEQTPPEKTSRITERLIWAGNTRAEKXHNLAGOCVRRHYDTAGVTRL	180		
Db	121	RYEGNTLPGRLLSVSEQVFNQESAKVTERFTWAGNTTSEKEYNSLGLCIRHYDTAGVTRL	180		
Qy	181	ESLSLTGTVLSQSSQLLIDTQEAANWTGDNETVMQNMLADDIYTTLTSTFDATGALLTQTD	240		
Db	181	MSQSLAGAMLQSQSHQLLAEGQEAANSGDDETVWQMLASEVYTTTQSTTNAIGALLTQTD	240		
Qy	241	KGNIQRLAYDVAGQNLNSWLTGKQTEQVILKSLTYSAAQOKLREEHGNVDVITEYSYEPE	300		
Db	241	KGNIQRLAYDIAGQLKGSWLTGKQSEQVIVKLSWSAAGHKLREEHNGVYVTEYSYEPE	300		
Qy	301	TORLIGIKTRR- - - - PSDTKVLQDLRYEYDPVGNVISIRNDAEATRFWHNOKWMPENTYT	356		
Db	301	TORLIGITTRAEQSGQARVQLQRLYKIDPVGNVISIHNDAEATRFWRNOKVEPENRYV	360		
Qy	357	YDSLYQLISATGREMANIGQSHQFPSPAL- - PSDNNYTYNTYTYTYDRGGLTKIOHS	414		
Db	361	YDSLYQLMSATGREMANIGQSNQLPSPVIVPPTDDSTYNTYTYTYDRGGLVQIRHS	420		
Qy	415	SPATQNNYTNITVSNRSNRAVLSTLTEDPAQVDALFDAGCHQNTLISGQNLNWNTRGEL	474		
Db	421	SPATQNSYTTDITVSSRSNRAVLSTLTTPTRVDALFDGCHQKGLIPGQNLWNIRGEL	480		
Qy	475	QOVTLVKRDKGANDREWRYSGDGRMLKINEQQASNNACQRTVYLPNLELTLQNST	534		
Db	481	QRVTPVSRN- -SSDSBWRYSSDGMRLLVSEQQTGNSTQVQRTVYLPGLLELTGVAD	538		
Qy	535	ATTEDLQVITVGEAGRAQVRVLHWESGKPEDIDNNQLRYSYDNLIGSSQLELDSEGOIIS	594		
Db	539	KTTEDLQVITVGEAGRAQVRVLHWESGKPTDIDNNQVRYSDNLIGSSQLELDSEGOILS	598		
Qy	595	EEYYPYGGTALWAARNQTEASYKTIYRYSKGERDATGLYYGYRYQYPWIGRWLSSDPAG	654		
Db	599	QEEYYPYGGTAIWAARNQTEASYKFIYRYSKGERDATGLYYGYRYQYPWIGRWLSADPAG	658		
Qy	655	TIDGLNLYRMVRNPNVTLLDPDGLMPT	681		
Db	659	TVDGLNLYRMVRNPNVTLLTDHDLAPS	685		

Search completed: February 16, 2006, 21:31:33
Job time : 100.189 secs

GenCore version 5.1.7
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: February 16, 2006, 21:43:14 ; Search time 33.2035 Seconds
(without alignments)
3670.210 Million cell updates/sec

Title: US-10-754-115-45
Perfect score: 7901.
Sequence: 1 MNSQDSITELSLPKGGGA.....WFTVNEBNDTAAEVKKVM 1474

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgm2_6/ptodata/1/iaa/5 COMB.pep.*
2: /cgm2_6/ptodata/1/iaa/6 COMB.pep.*
3: /cgm2_6/ptodata/1/iaa/H COMB.pep.*
4: /cgm2_6/ptodata/1/iaa/pCTUS COMB.pep.*
5: /cgm2_6/ptodata/1/iaa/RE COMB.pep.*
6: /cgm2_6/ptodata/1/iaa/backfiles.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6042	76.5	1481	2	US-09-251-645-14
2	6031.5	76.3	1476	2	US-09-817-514A-4
3	4581.5	58.0	1485	2	US-08-851-567B-32
4	160.5	2.0	1426	2	US-09-492-709A-340
5	155.5	2.0	658	2	US-09-252-991A-24910
6	155.5	2.0	2315	2	US-09-543-681A-5434
7	154	1.9	4630	2	US-09-091-609-2
8	154	1.9	5215	2	US-09-105-537-2
9	153.5	1.9	1028	2	US-09-543-681A-7181
10	144	1.8	1377	2	US-09-711-164-467
11	143	1.8	798	2	US-09-489-039A-10045
12	142.5	1.8	1183	1	US-08-447-031A-2
13	140	1.8	979	1	US-08-346-455B-38
14	140	1.8	979	2	US-08-977-221-38
15	140	1.8	979	2	US-09-483-821B-70
16	140	1.8	979	4	PCT-US95-06613-38
17	139.5	1.8	646	2	US-09-902-540-10353
18	139.5	1.8	1577	1	US-08-793-824-2
19	139	1.8	2200	2	US-09-796-575-2
20	138	1.7	1529	2	US-09-215-694-1
21	138	1.7	1529	1	US-10-109-310-1
22	137	1.7	2628	1	US-08-570-311-14
23	136.5	1.7	990	1	US-08-232-540-2
24	136.5	1.7	990	1	US-08-428-949A-2
25	136.5	1.7	990	1	US-08-428-948A-2
26	136.5	1.7	990	1	US-08-428-946-2
27	136.5	1.7	990	4	PCT-US95-04656-2

ALIGNMENTS

RESULT 1

US-09-251-645-14
; Sequence 14, Application US/09251645
; Patent No. 6281413
; GENERAL INFORMATION:
; APPLICANT: Kramer, Vance C.
; APPLICANT: Morgan, Michael K.
; APPLICANT: Anderson, Arne R.
; APPLICANT: Hart, Hope
; APPLICANT: Warren, Gregory W.
; APPLICANT: Dunn, Martha
; APPLICANT: Chen, Jeng S.
; TITLE OF INVENTION: NOVEL INSECTICIDAL TOXINS FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: CGC1963/A
; CURRENT APPLICATION NUMBER: US/09/251,645
; CURRENT FILING DATE: 1999-02-17
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 14
; LENGTH: 1481
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-251-645-14

Query Match 76.5%; Score 6042; DB 2; Length 1481;
Best Local Similarity 76.1%; Pred. No. 0;
Matches 1127; Conservative 113; Mismatches 226; Indels 14; Gaps 5;

Qy	1	MNSQDSITELSLPKGGGALTGMGEALTPTGPGMAALSPLPISACRGVAPAFTLN	60
Db	1	MNSQTSMTLSLPKGGGALTGMGEALTPTGPGMAALSPLPISACRGVAPAFTLN	60
Qy	61	SGAGNSPFLGWCNVMITRRTHFGVPHYDETDTFLGPEGEVLVA-----DQPRDS	114
Db	61	SGTNSPFLGWCNVMITRRTHFGVPHYDETDTFLGPEGEVLVAALNEAGQADIRSES	120
Qy	115	TLOGINLGAFTVTGYRSRLSHFSRLRYWPKTKTDFWLIIYSPDGOVHLLGKSPQAR	174
Db	121	SLOGINLGMFTVTGYRSRLSHFSRLRYWPKTKTDFWLIIYSPDGOVHLLGKSPQAR	180
Qy	175	ISNSQTSMTLSLPKGGGALTGMGEALTPTGPGMAALSPLPISACRGVAPAFTLN	234
Db	181	ISNPLNVMQTAQWLEASVSSHGQIYYQYRAEDTGCDEIHTHLQATQRYLHVY	240
Qy	235	GNRTASETLPGIDGSAFQADWLFYLVFDYDYSRNLKTPPAFTTSGWLCKRQDRFSRYE	294
Db	241	GNLTASEVFTLNGDDPLKSGWLFCLFYDYGRKNLSLSEMPFPKATSNWLCKRQDRFSRYE	300
Qy	295	YGFBIIRRLCRQVLMYHHLQALDLSKITEHNGPTLVSRILNLYDESAIASTLVFVRVGH	354

Sequence 8, Appli
Sequence 1, Appli
Sequence 4294, Ap
Sequence 2, Appli
Sequence 2, Appli
Sequence 36, Appli
Sequence 36, Appli
Sequence 36, Appli
Sequence 3, Appli
Sequence 23805, A
Sequence 4476, Ap
Sequence 6274, Ap
Sequence 3, Appli
Sequence 5, Appli
Sequence 21474, A
Sequence 14853, A

Qy	595	LPGFSQATFENPAQVVIADLDGSGPTDLIYVHTNRLDIFLANKSGNGFAPBVPVTLRPBGL	654
Db	601	LPGFSQASAFNPDVRHLADLDGSGPADLIYVHADRLDIFSNBSGNGFAKPFVTLSPDGL	660
Qy	655	RFQHTCQOMADVQGLGVASILLISVPHMSPHWRCDLTNNKMWLLMNNNMVGHVHTLRY	714
Db	661	RPDTCQGVADVQGLGVSVLISVPHMAPHWRCDLTNAKPWLLSETNNMGANHTLHY	720
Qy	715	RSSSQFWLDEKAALATTGQTPVCYLPPIHTLMQTEDEISGNKLVTTILRYARGAWDGR	774
Db	721	RSSVQFWLDEKAALATGQTPVCYLPFPVHTLMQTEDEISGNKLVTTILRYAHGAWDGR	780
Qy	775	ERERFGYGYEQTDSHQAQGNAPERPPALTKNWTATGLPVTDNALSTEYWR-DDQAFA	833
Db	781	ERERFGYGYEQTDSHQAQGNAPERPPALTAKSWATGLPVDNALSAGYWRGDKQAFA	840
Qy	834	GFSRPFVTTQDNKDVPVLTPEDDNSYWFNRALKGOLLRSYGLDSTNKHVYTVVTEFR	893
Db	841	GTFPRFTLWEGKDVLTPEDDHNLWYNALKGQPLRSYGLDGSAAQQIPIYVTTESR	900
Qy	894	SQVRLQHTDSRYPVLWSSVVEGRNHYERIASDPQCSQNTILSSDRFGQPLKOLSVQVP	953
Db	901	POVKQLQDQATVSPVLWASVVEGRSYHYERIIISDPQCNQDITLSSDLFGQPLKQVSVQVP	960
Qy	954	RROQPAINLPDPTLPDKLLANSYDDQORQLRLTYQSSWHHLTNVTVLGLPDSTSDI	1013
Db	961	RNNKPTNPPDPTLPDPTLFASSYDDQOQLRLTCROSSWHHLTGNELRVLGLPDGTRSDA	1020
Qy	1014	FTYGAENVPAGGLNLELLSDKNLSIADDKPREYLGQOKTAYTQGNQNTPLQOTPTROALIA	1073
Db	1021	FTYDAKVQPDVGLNLELTCAENLSIADDKPREVLNQORTYTTQKQKQPLKTPTROALIA	1080
Qy	1074	FTETTVFNOSTLSAFNGSIPSDKLSTLLEQAGYQQTNYLPPRTGEBKVMVAHHGYTDYGT	1133
Db	1081	FTETAVLTESILLSAFDGGITPDELPGILTQAGYQOEYPYLPFRTEGENKVMVARQGYTDYGT	1140
Qy	1134	AAQFWRPQKOSNTQLTGKIILINDANYCVVQTRDAAGLTTSKYDWRFLTPVOLTDIND	1193
Db	1141	EAQFWREVAQRNLSILTKMLKWDTHYCVITQTDAAGLTVSANYDWRFLTPQLTDIND	1200
Qy	1194	NOHLITLDALGRPTTLRFWGTENGKMTGYSSPEKASFPSPDVNAALTELKKPLPVAQCOV	1253
Db	1201	NVHLITLDALGRPVTVQRFWGTENGKMTGYSSPEKASFPSPDNDTDALTNGPLPVAQCLV	1260
Qy	1254	YAPESMWPLVSQKTFNRLABQDMQKLNYARIITTEDGRCTTLAYERVWQSQKALPQJLSLL	1313
Db	1261	YAPDSMWPLFSQETFNLTQESQETLRDSRIITEDWRICALTRRRMLQSKISTPLVKLL	1320
Qy	1314	NNGPRLPPHSLTLTTDRYDHDPEQIRBOQVVFSDGFGELLQAAARHEANGHARORNEGSL	1373
Db	1321	TNSIGLPPHNLTLTTDRYDRDSEQIRBOQVAFSDGFGRLLOASVRHEAGEAWORNOGSL	1380
Qy	1374	IINVOHTENRWAVTGRTEYDNKGOPIRTYQPYFLNDWRVYVNSDSAQEKEAYADTHVYDP	1433
Db	1381	VTKVNTKTIRWAVTGRTEYDNKGQIRTYQPYFLNDWRVYVSDISAR--KEAYADTHIYDP	1438
Qy	1434	IGREIKVITAKGWRPRTLFTFWFTVBNDENDTAAE	1468
Db	1439	IGREIRVITAKGWLROSOYFPFWFTVBNDENDTAAAD	1473

APPLICANT: Rocheleau, Thomas A.
 APPLICANT: Blackburn, Michael B.
 APPLICANT: Hey, Timothy D.
 APPLICANT: Merlo, Donald J.
 APPLICANT: Orr, Gregory L.
 APPLICANT: Roberts, Jean L.
 APPLICANT: Strickland, James A.
 APPLICANT: Guo, Lining
 APPLICANT: Ciche, Todd A.
 APPLICANT: Sukhapinda, Kitisri
 TITLE OF INVENTION: Insecticidal Protein Toxins
 NUMBER OF SEQUENCES: 88
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dow AgroSciences Patent Department
 STREET: 9330 Zionsville Road
 CITY: Indianapolis
 STATE: IN
 COUNTRY: US
 ZIP: 46268
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,567B
 FILING DATE: 05-MAY-1997
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/063,615
 FILING DATE: 18-MAY-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/395,497
 FILING DATE: 28-FEB-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/007,255
 FILING DATE: 06-NOV-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/608,423
 FILING DATE: 28-FEB-1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/705,484
 FILING DATE: 28-AUG-1996
 ATTORNEY/AGENT INFORMATION:
 NAME: Seay, Nicholas J
 REGISTRATION NUMBER: 273986
 REFERENCE/DOCKET NUMBER: 960296.93804
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 608-251-5000
 TELEFAX: 608-251-9166
 INFORMATION FOR SEQ ID NO: 32:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1485 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-851-567B-32

Query Match	58.0%;	Score 4581.5;	DB 2;	Length 1485;
Best Local Similarity	57.6%;	Pred. No. 0;		
Matches 858;	Conservative 224;	Mismatches 378;	Indels 29;	Gaps 11;

[illegible]

172 QARISNPQTTOTTAOWLLEASVSSRGEOIYVQRAEDDTGCEADEITHLQATAQRYLHI 231
 180 QACLANPONQOQTAQWLEETVTPAGHEVSYQVRAEAEHCDNEKTAHPWTAQRYLVQ 239
 232 VYTGNRATSETLPGLDGSAFSAQDWLFYLVDFYDGERSNKLTTPAFST-TGSMCLQRDRF 290
 240 VNYGNKPOASLFLVDNAPAPAESEWLFHLVFDHGERDTSLHTVPTWDAGTAQMSVRPDI 299
 291 SRBYEGEITRRLCROVLVYHHLQALDSKITHEINGPTLVSRLLINLYNDESAISTLVFVR 350
 300 SRBYEGEIVTRRLCQOVLVYHHLQALDSKITHEINGPTLVSRLLINLYNDESAISTLVFVR 359
 351 RVGHEQGNVUTLPPLELAVQDFSPRHHAHQWMDVLNANFNAIQRMOLVDLKGEGPLG 410
 360 QLSHESGRPVTPQPLELAWQRFDELEKIPTWQRFADLNFNSQORYQLVDLREBGLPGL 419
 411 YQKGAWYRSARQRLGIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDQGLDWVITGPG 470
 420 YQDGAWYKAPQREQDGSNAVYDKIAPLPTLNLQDNASLMDINGDQGLDWVITGPG 479
 471 LRGHVSQRPDGSWTRFTPLNALPVEYTHPRAQALADLMGAGLSDLVLIGPKSVRLYANTRD 530
 480 IRGHSQOPDQKWTFTFPIALPVEYHPHSIQFADLTGAGLSDLVLIGPKSVRLYANQNR 539
 531 GFAGKDVOSGDITLVPVCGADPKLVAFSDVLGSGOAHVVEVSATKVCWPNLGRGRFG 590
 540 GWRKEDVPOSTGILTLPVTGDARKLVAFSDVLGSGOAHVVEVSATKVCWPNLGRGRFG 599
 591 QPITLPGFSOPATFENPAQVYLADLDGSGPTDLIYVHTNRDLDFLNKSGNGFAPVTLRF 650
 600 QPLTSLGFSOPENSFNERFLADLDGSGPTDLIYVHTNRDLDFLNKSGNGFAPVTLRF 659
 651 PEGRLPHTCQLOQWADVQGLVASLILSVPHMSPHWRCDLTNWKPHLNMNMNMVHH 710
 660 PEGVQFQNTCQLOQWADVQGLVASLILSVPHMSPHWRCDLTNWKPHLNMNMNMVHH 719
 711 TLRVSSQFWLDEKAAALTTGQTPVCVLPFPHLTMOTETEDISGNKLVTLTVRYAGA 770
 720 TLYRSAQFWLDEKALQTVAGKSPACLPFPFPHLTMOTETEDISGNKLVTLTVRYAGA 779
 771 WGRERFRFGYVQVQSDSHQLAQGNAPERTPALTKNMYATGLPVIDNALSTYWR-DD 829
 780 WDKERERFRFGYVQVQSDSHQLAQGNAPERTPALTKNMYATGLPVIDNALSTYWR-DD 839
 830 QAFAGFSPRTTQ--DNKDVPLTPEDDNRVFNWALKQGLRSELYGLDDSTNKVVPY 887
 840 QAYSGFETRYTVDHTNQTDQAFETP-NETQRNMLTRALKGQLRLTELYGLDGTDKQTPY 898
 888 TVTEFRSQVRLQHTDSRYVPLASSVVEVRYHVERIASDPOCSNITLSSDRFGQPLKQ 947
 899 TVSESRYQVRSIPVWKETELSAWTALENKSYHYERIITDPOFSQSLQHDIFGQSLQS 958
 948 LSVQYPRQOPAINLYPDTLPDKLLANSYDDQQLRLTYQSSWHHLTNTTVRLGLPD 1007
 959 VDIAPWEREPANVPYPTLPETLFDSSYDDQQLRLTYQSSWHHLTNTTVRLGLPD 1018
 1008 STSDIPTGAENVAPAGLNLLELSDKNSLIADKPREYIQQOKTATYDQONTTLOPT 1067
 1019 AQRDVTYDRSKIPTREGISLEILLKDDGLADEKAAVYLGQQQTFYTAGQAEVLEKPT 1078
 1068 RQALIAFETTTVFNQSTLAFNGSIPSDKLSLTLLEQAGYQOQTNVLRPRTGEDKVVAAHG 1127
 1079 LQALVAFQETAMDDTSLQAGVIEQEQLNTALTOAGYQOQVAKLFNTRSESPWAARQ 1138
 1128 YTDYGTAAQFWRPQKQNTQITLWIDANYCVVQTRDAAGLTTSAKYDWRFLTPVQ 1187
 1139 YTDYGDAAQFWRPQKQNTQITLWIDANYCVVQTRDAAGLTTSAKYDWRFLTPVQ 1198
 1188 LTDINDNQHILTDALGRPTLRFWGTENGWMTYCSPEKASPSPPSDVNAALTELKPLP 1247
 1199 LTDINDNQHILTDALGRPTLRFWGTENGWMTYCSPEKASPSPPSDVNAALTELKPLP 1255

1248 VAQCOVYAPESWMPVLSQKTFNRL---AEQDWOKLYNARIITTEDGRICCTLAYRRWVQSQK 1304
 1256 VAQCLVAVDSWMPVLSQKTFNRL---AEQDWOKLYNARIITTEDGRICCTLAYRRWVQSQK 1315
 1305 AIPOLISLLANGPRLPHSLTLLTDYDHPDQEQIROQVVSFSGFGRLLQAAARHEAGMA 1364
 1316 LTIQLISLLASIPRLPHSLTLLTDYDHPDQEQIROQVVSFSGFGRLLQAAARHEAGMA 1375
 1365 ROENEGSLLINQ-----HTENRVAVTGRTEDNKGQPIRTVQVFLNDRWVVSND 1418
 1376 WQREKGLVNDANGVLSVAPTDRWAVSGRTEDNKGQPIRTVQVFLNDRWVVSND 1435
 1419 ROKEAYADTHVYDPIGRIKIVITAKGFRRTLTFTWFTVNEDENDTAA 1467
 1436 RD-DLFADTHLYDLGRIKIVITAKGFRRTLTFTWFTVNEDENDTAA 1482

RESULT 4
 US-09-492-709A-340
 ; Sequence 340, Application US/09492709A
 ; Patent No. 6720139
 ; GENERAL INFORMATION:
 ; APPLICANT: Zyskind, Judith
 ; APPLICANT: Ohlsen, Karl L.
 ; APPLICANT: Trawick, John
 ; APPLICANT: Forsyth, R. Alllyn
 ; APPLICANT: Freulich, Jamie M.
 ; APPLICANT: Carr, Grant J.
 ; APPLICANT: Yamamoto, Robert T.
 ; APPLICANT: Xu, H. Howard
 ; TITLE OF INVENTION: GENES IDENTIFIED AS REQUIRED FOR PROLIFERATION IN
 ; FILE REFERENCE: ELITEA 001A
 ; CURRENT APPLICATION NUMBER: US/09/492,709A
 ; CURRENT FILING DATE: 2000-01-27
 ; NUMBER OF SEQ ID NOS: 485
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 340
 ; LENGTH: 1426
 ; TYPE: PRT
 ; ORGANISM: E. Coli
 ; US-09-492-709A-340

Query Match 2.0%; Score 160.5; DB 2; Length 1426;
 Best Local Similarity 19.0%; Pred. No. 0.00045;
 Matches 255; Conservative 152; Mismatches 446; Indels 487; Gaps 65;

406 LPGLLYQDKGAWYRSARQRLGIGSDAVTWKMQPLSVIPSLQSNASLVLDINGDQGLDWV 465
 70 LPFILSRYSYRTKTPAPVGVFGP---WK--APSDIRLQLRDDGLILNDNGRSIHFE 124
 466 ITGPG-----LRGHSQRPDGSWTRFTPLNALPVEYTHPRAQALADLMGAGLSD 513
 125 PLLPGEAVYRSRSMVLVRGKKAQPDGH-----TLARLWGLALPPD 165
 514 LVILGPKSVRLYANTRDGFAGK-----KDVQSGDITLPGVADPKLVAFSDVLG-- 564
 166 IRL-----SPHLYLATNS--AOGFWILGWSERVPGAEVDVLPALPPVRLTGMADRGRT 219
 565 -----SGOAHVVEVSATKVCWPNLGRGRFGQIPIT 594
 220 LTVRREAAGDLAGEITGVTGAGREFRLVLTQQAAREARTS-----SLSSDSSRPLS 274
 595 LPGF--SOPATEFNP-----AQVYLDLDSGPTDL-----IVVHTNRDLDFLNK 638
 275 ASAFPDTLPGETYGPDRGIRLSAVLMLH--DPAPESLFAAPLVRYTTEAGELLAVYDRS 333
 639 GNGFAEPVTLRFPEGLRFDHTCQLOQWADVQGLVASLILSVPHMSPHWRCDLTNWKPWL 698
 334 NTQV-----RAFTYD-----AQHFGRWVHRYAGCPEN---RYRYDDTGR---- 370
 699 LNEKNNNMVHHTLRYRSSSQFWLDEKAAALTTGQTPVCVLPFPHLTMOTETEDISGN 758


```

Db 371 VVEQLNPAGLSVRYLYEQ-----DRITVDSLNREV-----LHT-----EGGAGL 411
Qy 759 KLVTTLRYARGAWDCRERFRGFGVEQTDHQLAQGNAPRTTPALTKNMYATGLPVID 818
Db 412 KRVVKELADGS-----VTRSGYDAAGRLTAQTDAGRRTEY--GLNVVS 454
Qy 819 NALS-----TEYRDD--QAPAGSPRETTWQDNKDVLPTPEDDNRWYFNALKG 867
Db 455 GDITDITTPDGRETFFYNDGNQLTAVVSPDGLSREYDEFGRLVSETSR-----SG 507
Qy 868 QLLRSELGLDDSTNKHVPYVTEFRSQVRLQLHTDSRVPVWSSVVSERNVHYERIASD 927
Db 508 ETVR---YRYDDA--HSELPATITDGTSTROM-----TWS-----RYGQLLAP 546
Qy 928 PQCSQNIIT--LSSDRFGQLKQLSVQPRQPAINLYPDTLPLDKLLANSYDDQQLRLT 986
Db 547 TDCSGYQRYEYDRFGQ-----MTAVHREBEGISLYRYDNRGRLTSVKAQGRETYE 599
Qy 987 Y-----QSSWHHLTNNTV-----RVGLPDSSTRS--DI 1013
Db 600 YNAAGDLTAVITPDGNSRSETQYDAMGKAVSTTQGLTRSMEDYDAAGRVISLTNENGSHV 659
Qy 1014 FTYGAEN--VPAGGLN-----LELLSPKNSLIA-----1039
Db 660 PSYDALDRLVQGGPDGRTQRYHYDLTKLQTSDEGLVILWYDESDRIHRTVTNGEPA 719
Qy 1040 -----DOKPR-----EYLGQOQT--AY 1054
Db 720 EQWQYDGHGWLTDISHLSEGRVAVHYGYDDKGRLTGECQIVENPETGELLWQHETKAY 779
Qy 1055 TD---GQNTTLPQTPTRO-----ALIAFTETTVFNOSTLS--AFNGSI 1092
Db 780 NEQGLANRVTPDLSPPVWELTVGSGYLACMKLGTPPLVEYTRDRILHRETIVRSFGSMAGSN 839
Qy 1093 PSDKLSLTLEAGYQOQTNLYLPTGCEKVAHAGYTDYGTAAQWRFQ--KQSTQLTG 1150
Db 840 AAYELTSTYTAGQLQSOHL-----NSLYDRDYGWSDGDLVRLSGPRQTRYEGYSATG 894
Qy 1151 KI---TLIWADANYCVVQTRDAAG-----LTSYAKYDWRFLTP 1185
Db 895 RLESVRTLAPDLDIRIPAT--DPAGNRLPDPELHPDSTLTWPDNRIAEADAHYVVRHDEY 953
Qy 1186 VOLTIN-----DNQHLITL-----DALGRPIT 1208
Db 954 GRLETKTDRIAGVIRTDDERTHYHDSQHLVFTYRIQHGEPLVESRYLYDPLGRMA 1013
Qy 1209 LRFWGTENGKWTGYSS---PEKASFSPSDVNAAILKPLPVAQCQ--VYAPESWMPVL 1263
Db 1014 KRVWRER--DLTGWMSLSRKPEVTWYGWGDRLTTVQ-----TDTTRIQTIVYEPGSGFTPLI 1068
Qy 1264 SOKTFNRLAEOQWQKLYNARIITEGRICTLAYRRWVQSKAIPOLISLLNNGPRLP--PH 1322
Db 1069 RVETENGREKA--QRSIAETLQEGS-----ENGHGVVEPA 1104
Qy 1323 SLTLTDRYHDPEQIQVQVFSDFGRLGLQAARHEAGMARQNE--GSLINVOHT 1380
Db 1105 ELVRLDLRL---EBEIRADRVSSESLAWLAQCGLTVEQ--LARQVEPYTPARKAHLVHC 1159
Qy 1381 ENR-----NAVYGRTEYD-----NKGQPIRTYQPYFNDWRYVNSDSARQ 1420
Db 1160 DHRGLPLALISDGNTAWS-----AEYDEWGNQLENPHHYVQPYRPLFGQOH-----DE 1209
Qy 1421 EKEAYADTH--VVDPI--GREI 1438
Db 1210 ESGLYNRRHYDPLQGRYI 1229

```

RESULT 5
 US-09-252-991A-24910
 ; Sequence 24910, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: ABRUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 24910
; LENGTH: 658
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-24910

Query Match 2.0%; Score 155.5; DB 2; Length 658;
Best Local Similarity 20.1%; Pred. No. 0.00031;
Matches 113; Conservative 61; Mismatches 155; Indels 233; Gaps 31;

Qy 356 QDGNVVTLPPLLEAYQDFSPRHHAHQPMQVLANFNAIQRMQLVDLKGEGL--PGLLYQ-- 412
Db 123 REGDNVTLRVNRNLAEDTS---IHWGIIILPANMDGVP---GLSFEIGIAPGGLYEVR 173
Qy 413 ---DKAWWYTRSQRLL-----GEIGSDA-----VTWEKMQPLSVIP 445
Db 174 FKVRQNGTYWYTHSHSGLEQAGVYGALVIDAREPEPFSYDRDYVLLSDWSEKQRIILA 233
Qy 446 SLQSNAS-----LVDINGDGLQDQWIT-----GP---GLRGY--- 474
Db 234 KLKQSDYNNFKRTKTVGDFIDDSANG---WAATLADRKKMAEMKMSFTDLADVSGYTYT 290
Qy 475 ---HSRQPDGSGWT-----RFTPLNAL-----PVEYT 497
Db 291 YLLNGQPPDGNGWTGLFRPGEKRLRFVNASAMSYFVDRIPGLKMTVVAADGQHVPEPSVD 350
Qy 498 HPRAQLADLMGAGLSDLVLGPKSVRLY-----ANTRDGPAGKDVQSGDITLTPVGAD 552
Db 351 ELRIATAETVD-----VIVEPGERAYTLFAQSMDRSGYARGTLALAEGL--LSAPVPTPD 403
Qy 553 PRKLVAFSVDLVSQAH-----LVVSATKVTWCWNLG-----585
Db 404 PRPLIGHDDMGGMGMDHGMCHGAATRPASENDHSKSGMDKMGMDHSGKMGMDMGMDH 463
Qy 586 ---RGRFGQPIITLPGFSQAPATE--FNPAQVYVLADLDGSGPTDLYIVHTNRDLIFLNK 637
Db 464 SKMAGMDHSRGMG--AMPQSHPASEDGNP---LVDMQTWTPTPKL-----ADPGLGL 512
Qy 638 SGNG-----FAPVTLRP--PEGLRFDHITCOLQADVQGLGVASLILSVPHMSPHHWCD 690
Db 513 RDNRRRVLTVAD--LRSRFADPDGREGPRTIELHTG-----HMEKFAWSFD 557
Qy 691 ---LTNNKRWLLNE-----MNNMGVHHTLRYRSSQFWLDEKAAALTTGTFVCYL 739
Db 558 GIKFSDAEPRLKTYGERLIRVLVNDTMMTH-----587
Qy 740 PPIHTLWQTEDEDEISGNKLV 761
Db 588 PIHLHGMW--SLEDE--QGNFLV 607

RESULT 6
US-09-543-681A-5434
; Sequence 5434, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706

```

; PRIOR FILING DATE: 1999-04-09

; NUMBER OF SEQ ID NOS: 8344

; SEQ ID NO 5434

; LENGTH: 2315

; TYPE: PRT

; ORGANISM: Proteus mirabilis

US-09-543-681A-5434

Query Match 2.0%; Score 155.5; DB 2; Length 2315;

Best Local Similarity 20.0%; Pred. No. 0.003;

Matches 313; Conservative 179; Mismatches 558; Indels 513; Gaps 81;

Qy 10 TELSLPQGGGAIATGGEALTPGDP-----GMAALSPL-PTISAGGY 51
 Db 662 TEVTVP--GGAAGDTLTLTKPDGSDTVEHTLTADVTAGKADVTIIPADKVTADGNY 719
 Qy 52 APAFTLNYSAGNSPFLGWCNVMTRRTHRGVPHYDSTDFTFLGPEGEVLVADQPR 111
 Db 720 --SVTAETIDPAGNTS-GQG-----KPTDFAVDTVAPSAPVLKAEDDDGVSVDLPT 767
 Qy 112 DESTLOGINLGATFTVGYRSLRSHFSRLEYWQPKTTGKTDFFWLIYSPD----- 161
 Db 768 D-----ANKGDTVEITFEDEKDKHTVLEKGDNGWTSPTPALI-----PDSNGDKATIPA 818
 Qy 162 -----GOVHLGKSPQARISNPSTOTTAQWLLLEASVSSRGEQIYYQYRAEDDTCCEAD 215
 Db 819 DNVDKNSVTCIAKDPGNSDPSTVTSKTDGVADAPVLFTIPEVA-----DGYANAD 870
 Qy 216 EITHLQA-----TAQRYLHV----- 232
 Db 871 ELKDGQAELVLPAGTVEGAEITLTVTRPKDTTETVTHVTVDKBAAGKVSVDIPKDAVQ 930
 Qy 233 -----YYGRTASETLPG-----LDGSPSQADWLFLYVFDYGRSNNLKTTPA 276
 Db 931 NGQNSVDVSLTQGNPAK---PGNKVDFAVDGQIPGDT-----GDGTVDTVP 976
 Qy 277 FSTTGMCLCRDRFSREYGEIETRLRCROVMVHHLQALDS---KITEHNGFTLVSR 333
 Db 977 VTIEPA---ADGNADELKDGVT-----QVTVPGSSAAGDTLTLTKPDGSDTIVEH 1027
 Qy 334 ILNYDESAIATLVFVRVGHGEQDGNV-----VTLP-----PLELAYQDFSRRHA 379
 Db 1028 TLTADEVTAGKADVTIPADKVTADGNTSVTAETIDPAGNTSGQKPTDFAVDTVAPS--- 1084
 Qy 380 HWQP-----MDVLNPNALQWQLV--DLKGEGLPGLLYQDKGAWWYSQAQLG 426
 Db 1085 --APVLKAEDDGSVVDLPTDANKGDTVEITFEDEKDKHTVLEKGDNGWTSPTPALIP 1142
 Qy 427 EIGSDAVT--WEKQPLSVIPLS-----OSNASLVDINGDQLDW-VITGPGLR-GYH 475
 Db 1143 DSNQDKATIPADNVKDNSEVTGIAKDPGNSDPSVTSTKTDGVADAPVLTIPEVADGYA 1202
 Qy 476 S--ORPDGSTRFTPLNALPVEYTHPRAQLADLWAGLSDLVLIGPKSVRLYAN--TRDG 531
 Db 1203 NADELKDG-----LQAEVLTPAGTIE--GAEITLTVTRPKDTTETVTHVTKDE 1249
 Qy 532 FAGK-----KQVQSG-----DITLPVPGADPRK---LVAFS-----DVLGSGQAH 570
 Db 1250 AAAGKVSVDIPKDAVQNSVDVSL--TQGNPNPAKGNKVDFAVDGQIPGDTDGDG---- 1304
 Qy 571 VEVSAKVTCPNPLGRG-----RFQPIITLPFSQAPATEFNPAPQVYLADLDGSGPTD 622
 Db 1305 -TVTTPVVTIPEATDGVNADELKDGQVTEVTVPGSSAAG---DTLTLTIKPDGS--TD 1358
 Qy 623 LI-----YVHTNRLDIFL-----NKSQN-----GFABPVTLRPPRGLRPHDTHCQLQM- 664
 Db 1359 TVEHTLTADVTAGKADVTIPADKVTADGNTSVTAETIDPAGNTSGQKPTDFAVDVTQIP 1418
 Qy 665 ADVQGLGVASL--ILSVPHMSPHWRCDLTNWKPLLNMNNMNVGHHTLYRYSQQFWL 722
 Db 1419 GDTGDGVDVDTTPVVTIPEAT-----DGVN-----ADELKDGVQTEVTVPGSGAA---- 1463
 Qy 723 DEKAAALTGTQTPVCYLPPF---IHTLWQTEDEISGNKLVTLTLRVARGAWDGREREFR 779

Db 1464 -----GDTLTLTKPDGSDTVEHTLTADVTAGKADVTIIPADKATPDG----- 1508
 Qy 780 GFGYVEQTDHSHQLAQNAPERTPPALTKWYATGLPVI-----DN-ALSTEVWRDQAFAGP 835
 Db 1509 --NYSVKAETIDPAGNTSGEGKATDFTVDTVAPSTPVLNAEDNGSVSVELPGD--ANKGD 1564
 Qy 836 SPRETTWQDNK-----DVPLTPEDDNRVWFNRALKQGLLR--SELY 875
 Db 1565 TVEI-TPEDSEKDKQVTMBKGDNGWTSDFNLIPDQGN-----NTAIPSDNVKDNSEVT 1619
 Qy 876 GL--DDSTNRKHVYVYTFEPRSVRRRLQHTDSRPVLMSSVVESSRNYHYERIASPOCSQ 933
 Db 1620 AIAKDPGNSAPATA-----MSKTDVLPVTSISVDTTSDVNDGDKISADASVSGE 1670
 Qy 934 ITL-----SSDRFGQPLKQLSVQYPRRQOPAINLYPDTL-----PD-KLLANSY 976
 Db 1671 ITEVPATIEDKDDTTGL-VYTVALDYVAAQDVTVTI---TLTNDAGHASAPDYSTLAGSQ 1726
 Qy 977 DDQORQL-----RLTYQQSSWHHL----- 995
 Db 1727 HDGKIALHGTGKVTYDGCATVTVVIPAGSKSVSFIYDPTMEANQNAFNAEGMEKVATI 1786
 Qy 996 ---TNNTVRVL-----GLPDSTRS-----DIFTYGAEN----- 1020
 Db 1787 TGTSNNAATVDTVNNAGASATGVIYDGNPISLRNLDDFTLKYSLSSVAENGDFGYTI 1846
 Qy 1021 -VPAGLNLELLSKNSLIADDPREYL-----GQOKTAYT-----DGQNTTPLQ 1064
 Db 1847 GVDGSKGKPKLTDDYNDTI-----YGVYQDGSETSSYSNLANSDQNGPDGKTGDNQ 1899
 Qy 1065 TPTRQALIAFTETTFVFNQSTLS-----APNGSIPSDKLSLTLEQ--AGYQQTNYLPFR 1115
 Db 1900 SITTVDLGAGDMLVRGNMLTNTRYVAGENDTFTMDGNALTARSMVAG---SYLFME 1955
 Qy 1116 TGEDKVVVAHHGYTDYG-----TAAQFWRPQKOSNTQLTKITLINDANYCVVVQT 1166
 Db 1956 SGNDVTVTKRTGVTNAGQIVLGSSTFTQ--GDANDQNNTLSGLLDL----- 2002
 Qy 1167 RDAAGLTTSKYDWRFLTPVQ-----LTDINDNHLITLDALGRITILRPMWGTENGK-WTG 1221
 Db 2003 --GSGTQDKSNMPNEXYLSVYQDGSNLSLGNNDN-----NIDAVTVDNTVTIYGSVSGEILGG 2056
 Qy 1222 YSS 1224
 Db 2057 YGS 2059

RESULT 7

US-09-091-609-2

; Sequence 2, Application US/09091609

; Patent No. 6600029

; GENERAL INFORMATION:

; APPLICANT: SHERMAN, DAVID H.

; APPLICANT: WILLIAMS, MARK D.

; APPLICANT: XUE, YONGQUAN

; TITLE OF INVENTION: METABOLIC ENGINEERING OF

; FILE OF INVENTION: POLYHYDROXYALKANOATE MONOMER SYNTHASES

; FILE REFERENCE: 600.297US2

; CURRENT APPLICATION NUMBER: US/09/091.609

; CURRENT FILING DATE: 1998-06-19

; EARLIER APPLICATION NUMBER: PCT/US96/20119

; EARLIER FILING DATE: 1996-12-18

; EARLIER APPLICATION NUMBER: 60/008,847

; EARLIER FILING DATE: 1995-12-19

; NUMBER OF SEQ ID NOS: 4

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 2

; LENGTH: 4630

; TYPE: PRT

; ORGANISM: Streptomyces venezuelae

US-09-091-609-2

Query Match 1.9%; Score 154; DB 2; Length 4630;
Best Local Similarity 21.0%; Pred. No. 0.014;
Matches 247; Conservative 123; Mismatches 426; Indels 380; Gaps 65;

QY 189 LEASVSSRGE---QIYYQYRAED---DTGCEAD-EITHLQATAQRYLHIVVYGNRTASE 241
DB 2026 IQVGVSGDGESPAQRTFGVSTPDSGDTGDDAPREWTRHVS-----VLGEQDPATES 2078
QY 242 TLPGLDGSAPSQADWLFVLVDYDGBSNNLKTTPAFSTTGSWLCRQDRFSRYEYGFERT 301
DB 2079 DHPGTGDD--GSAAM-----PPAAATATPLDGVYDLAEELGYG--- 2115
QY 302 RRLCRQVLMYHHLQALDLSKITEHNGPTLVSRLLIINYDESAIASTLVFVR--RVGHEQDN 359
DB 2116 -----GPAFOGLTGLWRD-----ADTLAEIRLPAQAQESAGL 2148
QY 360 VVTLPP--LELAYQDFSPRHHAHQPMQVLANFNAIQR-----WQVLDKGEGLPGLLYQ 412
DB 2149 FGVPALLDAALHPVILEGNSAAGACD--AUTDATDRIRLPPAWAGVTLHAEGATAL--- 2203
QY 413 DKGAWVRSQAQRLGIGSDAVTWKMQ-----PLSVIPSQSNASLVDINGD--GOLD--- 463
DB 2204 -----RVRIITGPTVTLRLTDTTGAPVAIVESITLRAVAKDRLGTAGRVDDAL 2254
QY 464 -----WVITG---PGLRGVHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLVL 516
DB 2255 FTVVWTGTGTPPAGRG-----AVEVE-----ELVDL--AGLGLDVE 2289
QY 517 IGPKSVRILYAN--TRDGFAGKDVVQSGDITLVPFAGDPKRLVAFSDVLGSGQ---AHLV 571
DB 2290 LGAADVLRADRWTLT-----GDPAAARTAVRRTLAIVQEFSEPRFPGSRILV 2338
QY 572 EVSATKVTWCNPLGRGRFGQPIIT-----LPGFSOPATEFNPAPQVYLADL--DGSQPTD-- 622
DB 2339 CVTRGAVAALP-----GEDVTSIATGELWGLVRSQAQSENPRGLFLDLGEGEGERDGA 2391
QY 623 -----LIYVHTNRDLIFLNKSGNGFAPVTLRF--PEGLRFDHTC 660
DB 2392 BELIRATAGDEPQLAARDGRLLAPRLARTAALESSEDTAGGAD-----RFGPDG-----T 2441
QY 661 QLOMADVOGLGVASLILSVPHMSPHW--RCDLTNKK-----PWLLENMNNWGVHTILRY 714
DB 2442 VLVTGCTGGLGA---LLARHLVHRVGRLLVSRRGADAPGAADGEDLAGLGAEVAF 2497
QY 715 RSSQFMLEKAAALTT--GQTPCVLPPEPIHTLQTE--TEDEISGNKLVTLR--YARG 769
DB 2498 AAADAADRESLARATVPAEHP---LTAHVHTAGVDDATVEALTPERDLAVLPKYDA 2554
QY 770 AWDGRE--REFRGFGYBQTDHQL-----AQGN--APERTPPALTKWYATGLPLVIDNAL 821
DB 2555 AWWLHETKDLRLDAFLVLFSSVSGIVGTAGQANYAAATGLDAAHRAATGLAA--TSL 2612
QY 822 STEYWRDQAFAGP--SPRFTTQONKDVPLTPEDDNRSYNPNRALKQGLRSELYGLDD 879
DB 2613 AMGLWDGTHGMGTGLGAADLARWSRAGITPLTP-----LOGALFDDAAVARD 2660
QY 880 STNKHPVYTVTFEFSQVRLQH--TDSRVPVLKSSVSRNHYHVERIASDPQCSQNITLS 937
DB 2661 AL--LVP-----AGLPTAHRGTGQPPALWRLVRAFRPRAARTAE-----A 2702
QY 938 SDRFQQLKQLSQVYP--RQQPAIINLYPDTLPDKLLANSYDDQQRQLRLTYQQSSWHHLT 996
DB 2703 ADTTGGWLSGLAAQSPSEERSSTAVTLVTGVAD----- 2735
QY 997 NNTVVLGLPDSSTRSDIPTYGAENVPAGLNLIELSDKNLSIADDPREYLGQQTATVD 1056
DB 2736 -----VLGHADSA-----AVGAER-----SFKD-----LGFDSLAGE 2763
QY 1057 GQN-----TTPLOTPTQALIAETETTFNQSTLSAFN-----GSIPSDKLST--TLEQAGY 1106
DB 2764 LENRLNAATGLRLPA-----TTVFDHPSFPAALASHLLAQVPLKEGTAATAVTVAE 2814
QY 1107 QCTNYLFPRTGEDKVVAAHGVTDYGTAAQF-----WRPQKQNTQLTG--KITLI 1155

DB 2815 RGASFGDRATDDDDPIAIV-----GMACRYPGGVSSPEDLWRLVAETDGAISEPVPNRG 2867
QY 1156 WD-----ANYCVVVVQTRDAAGLTTSYAKYDWRF--LTPVQVLTIDINDNQHLL--- 1197
DB 2868 WDLBSLYDPDPESKGTTC-----REGFLGAGDFDAAFGIGSPREALVMDPQQRLLLE 2922
QY 1198 ITLDALGR-----PITLRFWGTENGKMTGYSSPEKAS 1229
DB 2923 VSWEALERAGIDPSSLR--GSRGGVYVGAHGSYAS 2956

RESULT 8
US-09-105-537-2
; Sequence 2, Application US/09105537A
; Patent No. 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/105.537A
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 5215
; TYPE: PRT
; ORGANISM: Streptomyces venezuelae
US-09-105-537-2

Query Match 1.9%; Score 154; DB 2; Length 5215;
Best Local Similarity 21.0%; Pred. No. 0.017;
Matches 247; Conservative 123; Mismatches 426; Indels 380; Gaps 65;

QY 189 LEASVSSRGE---QIYYQYRAED---DTGCEAD-EITHLQATAQRYLHIVVYGNRTASE 241
DB 2026 IQVGVSGDGESPAQRTFGVSTPDSGDTGDDAPREWTRHVS-----VLGEQDPATES 2078
QY 242 TLPGLDGSAPSQADWLFVLVDYDGBSNNLKTTPAFSTTGSWLCRQDRFSRYEYGFERT 301
DB 2079 DHPGTGDD--GSAAM-----PPAAATATPLDGVYDLAEELGYG--- 2115
QY 302 RRLCRQVLMYHHLQALDLSKITEHNGPTLVSRLLIINYDESAIASTLVFVR--RVGHEQDN 359
DB 2116 -----GPAFOGLTGLWRD-----ADTLAEIRLPAQAQESAGL 2148
QY 360 VVTLPP--LELAYQDFSPRHHAHQPMQVLANFNAIQR-----WQVLDKGEGLPGLLYQ 412
DB 2149 FGVPALLDAALHPVILEGNSAAGACD--AUTDATDRIRLPPAWAGVTLHAEGATAL--- 2203
QY 413 DKGAWVRSQAQRLGIGSDAVTWKMQ-----PLSVIPSQSNASLVDINGD--GOLD--- 463
DB 2204 -----RVRIITGPTVTLRLTDTTGAPVAIVESITLRAVAKDRLGTAGRVDDAL 2254
QY 464 -----WVITG---PGLRGVHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMGAGLSLVL 516
DB 2255 FTVVWTGTGTPPAGRG-----AVEVE-----ELVDL--AGLGLDVE 2289
QY 517 IGPKSVRILYAN--TRDGFAGKDVVQSGDITLVPFAGDPKRLVAFSDVLGSGQ---AHLV 571
DB 2290 LGAADVLRADRWTLT-----GDPAAARTAVRRTLAIVQEFSEPRFPGSRILV 2338
QY 572 EVSATKVTWCNPLGRGRFGQPIIT-----LPGFSOPATEFNPAPQVYLADL--DGSQPTD-- 622
DB 2339 CVTRGAVAALP-----GEDVTSIATGELWGLVRSQAQSENPRGLFLDLGEGEGERDGA 2391
QY 623 -----LIYVHTNRDLIFLNKSGNGFAPVTLRF--PEGLRFDHTC 660
DB 2392 BELIRATAGDEPQLAARDGRLLAPRLARTAALESSEDTAGGAD-----RFGPDG-----T 2441

RESULT 10
 US-09-711-164-467
 ; Sequence 467, Application US/09711164
 ; Patent No. 6589738
 ; GENERAL INFORMATION:
 ; APPLICANT: Forsyth, R. Allyn
 ; APPLICANT: Ohlsen, Kari
 ; APPLICANT: Zyskind, Judith
 ; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERAPY
 ; FILE REFERENCE: ELTRA.008A
 ; CURRENT FILING DATE: 2000-11-09
 ; PRIOR FILING DATE: 1999-11-09
 ; NUMBER OF SEQ ID NOS: 469
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 467
 ; LENGTH: 1377
 ; TYPE: PRF
 ; ORGANISM: Escherichia coli
 US-09-711-164-467

Query Match 1.8%; Score 144; DB 2; Length 1377;
 Best Local Similarity 18.6%; Pred. No. 0.012; Mismatches 403; Indels 472; Gaps 62;
 Matches 230; Conservative 133;

QY 406 LPGLLYQDKGAWYRSQAORLGBIGSDAVTWKMQPLSVIPSLQSNASLVINDGQQLDWV 465
 DB 70 LPPILSRYSYRTKTPAPVGLSGPG--WK--MPADIRLQRLDNTLLSDNGRSLYFE 124
 QY 466 ITGPGLRGHSRPGSGWT-----RFTPL-NALPVEY-THPRAQLAD----- 505
 DB 125 HLFPGEDGY--SRSESLVLRGVAKLDEGHRLAALWQALPEELRLSPHYLATNSPOQP 182
 QY 506 --LMG-----AGLSDLVLICP-KSVRLYANTRDGPAGKDVQ-----SGDITLVPQAD 552
 DB 183 WLLGWCERVPDEALPAPLPYVLTGLVDRFGRTQTFHREAAAGEFSGEITVTDGAW 242
 QY 553 PR-KLVAFSDVLGSGOAHLEVS-ATKVTCPNLGRFGQPIITLPGFSQAPATEFNPAQV 610
 DB 243 RHFRVLITTOAQAAREARQAISGGTEPSAPFD-----TLPGYTE----- 282
 QY 611 YLADLDGSGPTDLIYVHTN--RLD-IFLNKSGNGFABPVTLFPPEGL-----RFDHTCQ 661
 DB 283 -----YGRDNGIRLSAVWLTH-----DP---EYPENLPAAPLVRYGWTGR 319
 QY 662 LQMA--DVQGLGVASLILSVPH-----MSPH-----HWRCDLTMKFWLLNEMNNMG 707
 DB 320 GELAVVYDRSGKQVRSFTYDDKYGRMVAHRHTGRPEIRYRYSDBG-----VTEQLNPAG 375
 QY 708 VHTLRY--RSSQFWLDEKAAALTGTGTPVCYLPFPHTLWQTEDEISGNKLVTL 764
 DB 376 LSYTYQEKDRITITDLSRREVLHTQGE-----AGLKRUVK 413
 QY 765 RYARGADGREREPFGYVE-QTDSHQAQGNAPERTPPALTKNMYATGLPVIDNALST 823
 DB 414 EHADGS--VTQSQFQAVGRLAQTD-----AAGRTTEYSPDVT--GLITRIITPDGRASA 465
 QY 824 EYWRDDQAPAGSPRFTWQDNKV-----PLTPEDNSRYWFWNPAKQLLR 871
 DB 466 FYNNHNLQTS-----ATGPDGLELRREYDELGRLIQETAPDGDITRY----- 510
 QY 872 SELYGLDSTNKHPVTVTFERSQVRRLQHTDSRPVLMSSVVESRNTHYERIASDPOCS 931
 DB 511 -----DNPHSDLP-CATE-----DATSRKMTWS-----RYGQLLSFTDCS 546
 QY 932 QNIT-LSSDRFGQ-----FLKQLSVQYPRRQOPAINLYPDTLPLKLLANSYDDQQRQ 982
 DB 547 GVTVRYDHRDFQMTAVHREGLSQYRAYDSRGLIAVK-----DTQGE 591
 QY 983 LRLTYQSSWHHLTNTNTRVRLGLPDSTRSDIFTYGAENVPAAGLNLELLSDKNLSIADK 1042
 DB 592 TRYEY-----NIAGDLTAVIAPDG 610

QY 1043 PREYLGQOKTAY-----TDQNTTPTLQTPTRQALIAFTETTVFNQSTLSAFNGSIPS-- 1094
 DB 611 SRN--GTQYDAWGKAVRTTQGLTRSMEDYDAAGRVIRLTSE-----NGSHTTFR 657
 QY 1095 -DKLSTTLEQAGY--QOTNYLFPFRG-----EDKVVVAHHGY-----TDYGTAAQF 1137
 DB 658 YVDLRLIOETFGDQRTQRYHDLTGKLIIRSEDEGLVTHWHYDEADRLTHRTVKGETAER 717
 QY 1138 WRPOKQ-----SNTQLTKG-----ITLIWD----- 1157
 DB 718 WQYDERGWLTDISHISEGHRVAVHYRDEKGLTGERQTVVHPQTEALLWQHETRAYNA 777
 QY 1158 ---ANYCVVQTRDAAGLTTSKAYDMRFLTPVQLTD-----INDNQHLITLDALGR-PI 1207
 DB 778 QGLANRCIPDSLPAVEWLT---YSGYLAGMKLGDTPLVEYTRDLRLHRETLRSFGRYEL 833
 QY 1208 TLRF-----WGTENGKWTGYSSEKA---SFSPSPDVNAALIEUK 1243
 DB 834 TTAYTPAGLOSOHLNLSLDRDYTM-NDNGELIRISSPRQTRSYSTTGRLTGVHHTA 892
 QY 1244 KPLPVAQCQVYAPESWMPVLSQKTFNRLAEQDWQKLYNARIITEDGRICTLAYRRWVQSQ 1303
 DB 893 ANLDIR-----IPYATDPAGNRLPD----- 912
 QY 1304 KAIPOLISLANNPRLPPHS-LTLTLD-RYDHPDQIQVQVFDGFRLLQAAARHEA 1361
 DB 913 -----PELHPDSTLSMMPDNRIARDAHYLYR-----DRHGLRTEKTDLIPE 954
 QY 1362 GMARQNEGSLIINVQHTENRWATGTEVDNKGQPIRTYQPYFLNDWRYVNSDSAROE 1421
 DB 955 GVITDDE-----RTHRYHYDSQHRLVHYTRTOYE 984
 QY 1422 KEAVADTHVYPIGRIKIVITAKGWFRTLTFTMFTVN 1459
 DB 985 EPLVESRILYDPLGR---VAKRVWRERDLTGWMSLS 1019

RESULT 11
 US-09-489-039A-10045
 ; Sequence 10045, Application US/09489039A
 ; Patent No. 6610836
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary Breton et. al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
 ; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: 2709.2004001
 ; CURRENT FILING DATE: 2000-01-27
 ; PRIOR FILING DATE: 1999-01-29
 ; NUMBER OF SEQ ID NOS: 14342
 ; SEQ ID NO 10045
 ; LENGTH: 798
 ; TYPE: PRF
 ; ORGANISM: Klebsiella pneumoniae
 US-09-489-039A-10045

Query Match 1.8%; Score 143; DB 2; Length 798;
 Best Local Similarity 20.9%; Pred. No. 0.0056;
 Matches 137; Conservative 81; Mismatches 232; Indels 206; Gaps 34;

QY 18 GGAITGMEALTPCTGMAALSLPLPISAGRGYAPAFATLNYNSGAGNSPFLGWDG--- 74
 DB 123 GYNIRGMSNRVGDVQIAQ-----PNATGRGV-----GRAGLNTFGIGRDYIDP 169
 QY 75 -----NMTIRRTHEGV-PHYDET----- 94
 DB 170 YMGSDVIQSGATSTETANSAGNVSPRPSADYLPFGKTSFAGSYGSDSADRSWHN 229
 QY 95 --TFGLPGEVL--VVADQPRDESTLQGINLGATFTVGYRSLRLESH-FSRLEYWQP--- 146
 DB 230 GVTVAGGD-EFLRGILVYRRDRDQETEN-NSG---TVDAYPANMHSDAFLASGIWQPNDE 284

QY 147 -KTTGKTPWLLYSPDGOVH--LLGKSPQARISNPSTOTTAQMLLEAS-----VS 194
Db 285 HKLTSTFDY-----YHKTNNHTYDWDSSNGSTIGTANQTSQTRWGLSLKDDWTPMNDYLD 341
QY 195 SRGEIYYQYRAEDDTGCEADIEITHLQATQRYLHIVYGNRTA-SETPLGLDGSAPSQ 253
Db 342 SVSYKIYYQTEAHDWTYMPDSVTRKMQTVNSNY-DTDTWGLQTLAKTLGRHDLA--- 397
QY 254 ADWLFLYVDFYGERNNLKTTPAPSTGTSWLCRODRFSRYEYGFETRRLCROVLMYHH 313
Db 398 -----GFNASTSKTORPFS----- 411
QY 314 LQALDSKITEHNGPTLVSRLLINVDSEIASTLVFVR-RVCGHEODG-NVVTLPPLLEYAQ 371
Db 412 -----QSPI-----PSVSEIHOPEADRSYTLGGFVQDKINFDFLDSDHNFVIGRVVHQ 462
QY 372 DFPSPRHAWQPMVDLANFNAIQRWLVDLKGE-----LPGLLYQ-----D 413
Db 463 STKENLS-----DLAANSSVLSESSVANLYGKNSDTQVLSLTFQYDLTPELMTYLYQY 517
QY 414 KGAWYRSARQLG--EIGSDAVTWKMQPLSVIPSLQSNASLVNDINGDGLDWITGPEL 471
Db 518 RGAQFPNASQLYGSWNLGS---SYAGSOYALI-----GNTDLKTETSD-NLEWGLKGEVT 569
QY 472 RG-----YHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMAGLSDLVLI--GPKS 521
Db 570 EGITLRTALFYNSYKNFIATRYTRANN-PCQFTNVPSNIYIYOENRDKAYIYGGEIS 628
QY 522 VR-----LYANTRDGFAGKGDVQV-QSGDITLVPVPGADPRKL---VAFSD 561
Db 629 TKFNGTWFQVDGLSATLALGYSEKSKSYSGDKYVLDLSDVAPMKAIVGVANDD 684

RESULT 12

US-08-447-031A-2
; Sequence 2, Application US/08447031A
; Patent No. 5851794
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: HOOK, Magnus
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: PATRI, Joseph
; APPLICANT: SIGNAS, Christer
; APPLICANT: SWITALSKI, Lech
; TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
; TITLE OF INVENTION: ITS PREPARATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,031A
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/861,804
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE91/00707
; FILING DATE: 22-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9003374-7

; FILING DATE: 22-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 012889-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-447-031A-2

Query Match 1.8%; Score 142.5; DB 1; Length 1183;
Best Local Similarity 17.5%; Pred. No. 0.013;
Matches 202; Conservative 180; Mismatches 419; Indels 355; Gaps 51;

QY 505 DLMGAGLSDLVLI-----GPKSVRLYANTRDGFAGKGDVQSGD-ITLVPVPGADPRKL 556
Db 32 DISSTNVTDLTVPSPKIEDGGKTVKMTFDDKNG-----KIQNGDMIKVAMWPTSGTVKI 85
QY 557 VAFS-----DVLGS--GQAHLVEVSATKVTCPNLRGRFGQPI-----TLPGFSQ----- 600
Db 86 EGYSKTIVPLTVKGEVQCAVITPDGAT-IT-----FNDKVKELSDVSGFAEFVQG 135
QY 601 -PATEFNPQVYLADLDSGSGFTDLIYVHTNRL-----DIFLNKSGNGFAEPVT-LRF----- 650
Db 136 RNLTOTNTSDKVATITSGNKSTNVTVKSEAGTSVFFYKTDGMDLPEDTTHVWFNLIN 195
QY 651 --PEGLRFDFTCOLQADVOGLGVASLILSV--PHMSPHHWRCDLTNKPKWL-----LN 700
Db 196 NEKSYVSKDITIKDQIQGGQOLDLSTLININVTGHSNYSGQSAITDFEKAPGSKITVD 255
QY 701 EMNNMVGHHLTRYRSSSQFWLDEKAAALATGTQPPVCVLPFPIHTLWQTETEDISGNKL 760
Db 256 NTKNTIDVTIPOGYSYNSFSINYKTKITNEQQKEFVN---NSQAWQEHGKEEVNGSKF 312
QY 761 VTLRYARGAWDGRERFRGFGVVEQTDSSHQLAGNAPERTPPA-----LTQNWYATGLP 815
Db 313 NHTV-HNINANAGIEGTVKELKVLKQDK-----DTRAPIANVAFKLSKK---DGSV 360
QY 816 VIDNALSTEWDRDQAFAGP-----SPRFTTWQDNKQVPLTPEDDNSRYWF 861
Db 361 VKDNQKEIETDANGIANIKALPSGDYILKIEAPRPYTFDKDKEYPFTWKQTDNQGIF 420
QY 862 NRALKGQLRS-----ELYGLDDSTNKHVPYTVTFEFSQVRL 899
Db 421 TTENAKAIEKTRKDVSAQKWEGTKVKPTIYFKLYKQDDNQ-----TTPVDKAEIKKL 475
QY 900 QHTDSRYPVLWSSVVESRNYHYERIAADPOCSQNTLSSDRFGQPLKOLSVOYPRPOPA 959
Db 476 E--DGTTKVTWSNLP-----NDKNGKAIKYLKVENAQQE-- 509
QY 960 INLYPDTLPD-----KLLANS-----YDQQRQLRLTVQSSSMHLLTN 997
Db 510 -----DTPTEGYTKKNGLVVVTNTEKPIETTSISGEKVDKQNGQDGRPEKVSNNLLAN 564
QY 998 -NTRVRLGLPDSTRSDIFTYGAENVPA--GGLNLELLSDKNLSLIADDPREYL----- 1047
Db 565 GEKVTL---DVTSETNWKYEFKDLPKYDEGKKLEY-----TVTEDHVVDYTTDINGTT 615
QY 1048 -----GQOKTAYT-----DQNTTLPQPTTQALIAFTTETTFNOS-----TLS 1086
Db 616 ITNKYTPGETSATVTKWDDNNNDGKRPTKVELYQDGKATGKTAILESNNNWHTWT 675
QY 1087 AFNGSIPEDKLSLTLEQ-----AGYQQ-----TNYLFPRT---GDKVWVAHHG 1127
Db 676 GLDEKAKGQVQKYTVEELTKVKGYTTHVDNNDMGNLIVTNKYTPETTSISGEKVV----- 730
QY 1128 YTDYGTAAQFWRPQKQNTQLTGKITLIWDANYCVVVQTRDAAGLTTSKAYDMRFL----- 1183

Db 731 --DDKNDQGRPEKVS-----VNLADGE---KVKTLDVTS-ETNWKYEFKDLPKYD 777
Qy 1184 --TPVQLTDINDNOHLITLIDALGRPITLRFWGTENGKMTGYSSPEKASFSPSDVNAAE 1241
Db 778 EGKLEYVTEDHVXDYTTDINGTITNKY-----TPGETSAIV- 816
Qy 1242 LKKPLPVAQCQVYAPESWMPVLQKTFNRLAEOQWOKLYNARIITEDGRICTLAYRRWQ 1301
Db 817 -----TKNWDNNN-----QDGKRPTEIKVELYQ 840
Qy 1302 SOKAIPQLISLIANGPRLPHSLTLTDRYDHPQOIRQOVVSDGFGRLLOAAARHEA 1361
Db 841 DKGATGK-TAILNES-----NNWTHWTGLD---EKAGQQVKYTVBELTKVGYTHV- 890
Qy 1362 GMAQRNEDGSLIINVQHTENRWAVTGRTEYDNK-----GQPIRTYQ 1403
Db 891 -----DNDMGNLIVTNKYTPETTSISGEKVWDDKNDQGRPEKVSNNLANGEKVKTL 946
Qy 1404 PYFLNDWRYVNSDSARQEK-----EAYADTHYDPIGREIKVITAKGWFRTLTFT 1454
Db 947 VTSETNWKYEPKDLPEKDEGKIEYVTEDHVKDVTDTDINGTITNKYTPGETSATVTKN 1006
Qy 1455 WFTVNEDENDTAAYK 1470
Db 1007 WDDNNQDGRKRPTEIK 1022

RESULT 13

US-08-346-455B-38
; Sequence 38, Application US/08346455B
; Patent No. 5731167
; GENERAL INFORMATION:
; APPLICANT: UNITED STATES OF AMERICA; DEPT.
; APPLICANT: OF HEALTH AND HUMAN SERVICES
; TITLE OF INVENTION: MOTILITY STIMULATING
; TITLE OF INVENTION: PROTEIN USEFUL IN CANCER DIAGNOSIS AND
; TITLE OF INVENTION: THERAPY
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/346,455B
; FILING DATE: 28-NOV-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06613
; FILING DATE: 24-MAY-1995
; APPLICATION NUMBER: 08/249,182
; FILING DATE: 25-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/822,043
; FILING DATE: 17-JAN-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: DOROTHY R. AUTH
; REGISTRATION NUMBER: 36,434
; REFERENCE/DOCKET NUMBER: 2026-4149PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:

LENGTH: 979
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: Unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: No
; ORIGINAL SOURCE:
; ORGANISM: Human
; STRAIN:
; INDIVIDUAL ISOLATE:
; DEVELOPMENTAL STAGE:
; HAPLOTYPE:
; TISSUE TYPE: Liver
; CELL TYPE:
; CELL LINE:
; ORGANELLE:
; FEATURE:
; NAME/KEY:
; LOCATION:
; IDENTIFICATION METHOD:
; OTHER INFORMATION: putative autotoxin
; OTHER INFORMATION: protein sequence from human liver
US-08-346-455B-38
Query Match 1.8%; Score 140; DB 1; Length 979;
Best Local Similarity 18.9%; Pred. No. 0.015;
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;
Qy 418 WYSAORLGEIGSDAVTWKMQPLSVIPSLQSNASLVSDINGDGLDWI-----TGPGLRG 473
Db 261 WITATKQGE-----SWN-----ILLVCHSRAELIT-----LQWTLDPHERPSVYA 305
Qy 474 YHSQRPDGSWTRFTPLNALPVEYTHRAQLADLMA-----GLSDLVLIQPKSVRLYANTRD 530
Db 306 FYSEQPDFSGHKMPFGP---EMTNPLREMHKIVGQMDGLKQLKLRVCNV-IFVETMD 361
Qy 531 GFAK-----GKDVQSGDITLPVGA-----DPRKLVAFSDVLGSGQAH 569
Db 362 GRCHMYRTEFLSNYLTNVDDITL-VPGTLGRIRSKFSNNAKYDPRKAITA----- 409
Qy 570 LVEVSATKVTCPNPLGRGRFGQITLPGFSOPATEFNEPAQVVLADLDGSGPTDLYVHTN 629
Db 410 -----NLTC-----KKPDQHFKE---YLKQ---HLPKRLHYANNR 438
Qy 630 R-----LDIFLNKSGNGFAEPVTLFPEGLRPFDHTCQLQMAADVQGLG 671
Db 439 RIEDIHLLVERRHVHVARPKLDVYKKPSGNASRETT-----AFDNKVNMQTVFVGVG 491
Qy 672 ----VASLILSVPHMSPHWRCDLTNNKPKWLLNEMNNMNGVHHILYRSSSQFWLDEKAA 727
Db 492 PTFKYTKVPPFENIELYNVNCDDLGLKP-----APNNGTHGSLNHLRLTNTFRPTMPE 545
Qy 728 ALTTGQTP-VCYLPFP-----IHTLMOTETEDETSGNKLVT 763
Db 546 EVTRPNYPGIMYLOSDFDLGCTCDDKVEPKNKLDELNKLRLHKGSTEEHLLYGBRPAVL 605
Qy 764 LR-----YARGAWDGRERERFGYVEQTDHQLAOGNAPE-----RTPPALTK 807
Db 606 YRTRYDILYHTDPESGYSEIEFLMPLWTSYTSYKQAEVSSVPDHLTSCVPRDVRVSPFSQ 665
Qy 808 NYATGLPVIDNALSTEWRRDDQAFAG-----SP--RFTTWQDNKDVPLTDEDDNS 857
Db 666 NCLA-----YKNDKQMSYGFPLFPFYLSSSSPEAKYDAPLVNTVMVPMYPAFKRV 712
Qy 858 RYWFNRAL-----KGOLLESELYGLDDSTNKHVPYTVTFERSQVRLQHTD 903
Db 713 WNYFORVLVKKYASERNGVNVISGIFDYDVGDLHDTEDKIKQYV-----EG 759
Qy 904 SRYPVLMSSVVESRNYHYERIASPPQCSQNTLSSDREGQPLKQLSVQVPRRQOPAINLY 963
Db 760 SSIPV-----PTHYYSIITS-----CLDTPQADKCDGPLSVSSSFLPHR----- 799
Qy 964 PDTLPDKLLANSYDDQORQLRLTYQSSW---HLLTNTNTRVVLGLPDPSTRSDIFTYGAENV 1021

FILE REFERENCE: 2026-4149US4
CURRENT APPLICATION NUMBER: US/09/483,831B
CURRENT FILING DATE: 2000-01-17
PRIOR APPLICATION NUMBER: 07/822,043
PRIOR FILING DATE: 1992-01-17
PRIOR APPLICATION NUMBER: 08/249,182
PRIOR FILING DATE: 1994-05-25
PRIOR APPLICATION NUMBER: 08/346,455
PRIOR FILING DATE: 1994-11-28
PRIOR APPLICATION NUMBER: 08/977,221
PRIOR FILING DATE: 1997-11-24
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 70
LENGTH: 979
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Putative autotoxin protein sequence from human
OTHER INFORMATION: liver.
NAME/KEY: VARIANT
LOCATION: (860)
OTHER INFORMATION: Xaa at positions: 860, 889, 905, 911, 927, 937,
OTHER INFORMATION: 944, 950, 954, 967, and 975 represents an unknown
OTHER INFORMATION: or other amino acid.
US-09-483-831B-70

Query Match 1.8%; Score 140; DB 2; Length 979;
Best Local Similarity 18.9%; Pred. No. 0.015;
Matches 139; Conservative 83; Mismatches 258; Indels 256; Gaps 34;

QY 418 WYSAQRLEIGSDAVTWKMQPLSVIPSLQSNASLVINGDGLDWI-----TGPGLRG 473
DB 261 WITATKQGE-----SWN-----ILLVCCHPSAEILTI-----LQMLTLPDHERPSVYA 305
QY 474 YHSQRPDGSWTRFTPLNALPVEYTHPRAQLADLMCA---GLSDLVLIGPKSVRLYANTRD 530
DB 306 FYSEQPDFSGHKMPFGP---EMTNPLREMHKIVGQLMDGLKQLKLRGVNV-IFVETMD 361
QY 531 GFPAK-----GKDVVQSGDITLPVPGA-----DPRKLVAFSDVLGSGQAH 569
DB 362 GRCHMYRTEFLSNLTNTVDDITL-VPGTLGRIRKFSNNAKYDPAIIA----- 409
QY 570 LVEVSATKVTCPNIGRFGQPIITLPGFSQATEFNPQAQVYLADLDGSGPTDLIYVHTN 629
DB 410 -----NLTC-----KKPDQHKP---YLKQ---HLPKRLHYANNR 438
QY 630 R-----LDIFLNKSGNGFAEPVTLRPEGLRFDHTCQLQOMADVQGLG 671
DB 439 RIEDIHLIVERHWHVARKPLDVIYKPKSGNAFSRETT-----AFDNKVNMQTVFVGYG 491
QY 672 ---VASILSVPHMSPHHWRCDLTNMKPLLNNMNMVGHHTLRYRSSQFWLDEKAA 727
DB 492 PTFKTKVPPFENIELYNMCDLGLRP-----APNNGTHGSLNHLRLRTNFRPTWPE 545
QY 728 ALTTGQTP-VCYLPFPF-----IHTLMQTEDEISGNKLVTT 763
DB 546 EVTRYNYFGIMYLOQDFDLGCTCDDKVEPKNKLDELNKLHTKSGSTEERHLLYGRPAVL 605
QY 764 LR-----YARGAWDGRERFRFGYVQGTSHQLAQNABE-----RTPPALTK 807
DB 606 YRTRYDILYHTDFESGYSEIFLPLMTSVTSKQAEVSVDPDLTSCVRPDPVRVSPFSQ 665
QY 808 NWTATGLPVIDNALSTYWRDDQAFG-----SP--RFTTWQDNKDVPPLTPEDDNS 857
DB 666 NCLA-----YNDKQMSYGFLLPFPYLSLSSPEAKYDAFLVTNMVPMYPAFKRV 712
QY 858 RYWFNRAL-----KGQLLRSELYGLDDSTNKXVPYTVTFEFSQVRRLOHTD 903
DB 713 WNYFORVLKKYASERNGVNIVSGIFDYDYDGLDHTEDKIQYV-----EG 759
QY 904 SRYPVLWSSVVESRNYHYERIASDPQCSQNTLSSDRFGQPLKQLSVQYPRRQPAINLY 963

Db 760 SSIPV-----PTHYYSIITS-----CLDFTQPADKCDGFLSVSSFILPHR----- 799
QY 964 PDTLPDKLLANSYDDQQRQLRLTYQSSW--HHLTNTNTRVRLGLPDSTRSDIFTYGAENV 1021
Db 800 FD---NEESSCNSSD-----ESKWVEELMOMHTARVRDIEHLTSLDFFRKTSSY 846
QY 1022 PAGGLNLELLSDKNSL 1037
Db 847 P-----EILTILTYL 856

Search completed: February 16, 2006, 21:46:34
Job time : 40.2035 secs

THIS PAGE BLANK (USPTO)
